

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2005, 19:14:15 ; Search time 166 Seconds
(without alignments)
1311.723 Million cell updates/sec

Title: US-09-936-759-6
Perfect score: 3001
Sequence: 1 MVRPPQRMKKRFILNLGVWN.....TRDRQPKLVAVHLRLMSEV 563

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

1: _geneseqp1980s:*

2: _geneseqp1990s:*

3: _geneseqp2000s:*

4: _geneseqp2001s:*

5: _geneseqp2002s:*

6: _geneseqp2003as:*

7: _geneseqp2003bs:*

8: _geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2999	99.9	563	3 AAB28406	Aab28406 Thermotog
2	2999	99.9	563	8 ADN20342	Adn20342 Bacterial
3	932	31.1	602	2 AAW93820	Aaw93820 Bacillus
4	929	31.0	602	2 AAW93825	Aaw93825 Bacillus
5	929	31.0	602	2 AAW93822	Aaw93822 Bacillus
6	929	31.0	602	2 AAW93826	Aaw93826 Bacillus
7	929	31.0	602	3 AAB28402	Aab28402 Staphyloc
8	929	31.0	618	2 AAW93821	Aaw93821 Bacillus
9	925	30.8	615	3 AAB28408	Aab28408 Codon-opt
10	920	30.7	602	3 AAB28409	Aab28409 Salmonell
11	919.5	30.6	598	5 AAM48998	Aam48998 L. gasseri
12	905	30.2	656	8 ADQ89612	Adq89612 Antagonis
13	902.5	30.1	648	7 ADF28924	Adf28924 Murine be
14	899.5	30.0	648	4 AAE02444	Aae02444 Murine be
15	897	29.9	643	4 ABB70164	Abb70164 Drosophil
16	894.5	29.8	648	7 ADD45479	Add45479 Rat Prote
17	894.5	29.8	648	7 ADE57446	Ade57446 Rat Prote
18	882	29.4	1010	3 AAY68840	Aay68840 Fusion pr
19	879	29.3	603	2 AAW93827	Aaw93827 E. coli G
20	879	29.3	603	2 AAW93824	Aaw93824 Human GUS
21	879	29.3	603	3 AAB28431	Aab28431 Human bet
22	879	29.3	603	5 ABB84108	Abb84108 GUS prote
23	879	29.3	603	6 ABP96657	Abp96657 E. coli b
24	879	29.3	603	6 ABJ19649	Abj19649 Artificia
25	879	29.3	603	8 ADN18189	Adn18189 Bacterial

26	877	29.2	618	7 ADL01666	Adl01666 Modified
27	875.5	29.2	602	2 AAR43387	Aar43387 Beta-gluc
28	875	29.2	604	5 ABB84107	Abb84107 GUS prote
29	875	29.2	604	7 ADD27986	Add27986 Beta-gluc
30	875	29.2	659	7 ADD27989	Add27989 Oleosin/b
31	875	29.2	850	7 ADD27991	Add27991 Caleosin/
32	872.5	29.1	602	2 AAW42429	Aaw42429 Escherich
33	872.5	29.1	607	8 ADS26258	Ads26258 Bacterial
34	872.5	29.1	711	6 ABR83626	Abrr83626 SUMO-beta
35	872.5	29.1	1242	5 ABB81108	Abb81108 LUC-U3'-U
36	872.5	29.1	1242	6 ABB84637	Abb84637 LUC-U3'-U
37	870	29.0	832	2 AAW04302	Aaw04302 Antibody/
38	869.5	29.0	602	1 AAP82948	Aap82948 Beta-gluc
39	865.5	28.8	613	2 AAW93823	Aaw93823 E. coli G
40	865.5	28.8	613	2 AAW93828	Aaw93828 Human GUS
41	865.5	28.8	613	3 AAB28407	Aab28407 Escherich
42	865.5	28.8	633	4 AAB62276	Aab62276 Mutant he
43	865.5	28.8	633	4 AAB62271	Aab62271 Heavy cha
44	865.5	28.8	651	4 AAE02443	Aae02443 Human bet
45	865.5	28.8	651	7 ADD45481	Add45481 Human Pro

ALIGNMENTS

RESULT 1

AAB28406

ID AAB28406 standard; protein; 563 AA.

XX AAB28406;

AC 26-JAN-2001 (first entry)

XX

DT Thermotoga maritima beta-glucuronidase.

XX

DE Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;

XX Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;

KW transgenic insect; marker; glucuronide detoxification.

KM

XX Thermotoga maritima.

OS

XX WO20005533-A1.

PN

XX 21-SEP-2000.

PD

XX 16-MAR-2000; 2000WO-US0071107.

PF

XX 17-MAR-1999; 99US-00270957.

PR (CAMB-) CAMBIA BIOSYSTEMS LLC.

PA Jefferson RA, Mayer JE;

XX

PI WPI; 2000-647075/62.

XX N-PSDB; AAA07937.

DR

XX Novel microbial beta-glucuronidase genes and gene products used as

PT reporter/effector molecule, as diagnostic tool, in positive selection, to

PT target molecules to specific cells and to detect and track linked genes.

XX

PS Claim 3; Fig 5B; 116pp; English.

XX

CC The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS

CC genes were obtained from six different genera: Enterobacter/Salmonella,

CC Pseudomonas, Salmonella, Staphylococcus and Thermotoga. Microbial GUS can

CC be used as a reporter/effector molecule for transgenic constructions and

CC in vitro diagnostic applications. It may also be used to generate

CC sentinel plants that serve as bioindicators of environmental status. It

CC may be used to generate transgenic insects for tracking insect

CC populations or to facilitate the development of a bioassay for compounds

CC that affect molecules critical for insect development (e.g. juvenile

CC hormone). Secreted GUS may also serve as a marker for beneficial fungi

CC destined for release into the environment. In animal systems, secreted

CC GUS may be used to achieve extracellular detoxification of glucuronides
CC (e.g. toxin glucuronide) and to examine conjugation patterns of
CC glucuronides. Microbial GUS may also be used in traditional medical
CC diagnostic assays, for drug testing, pharmacokinetic studies,
CC bioavailability studies, diagnosis of diseases and syndromes, following
CC progression of disease or its response to therapy. Microbial GUS has
CC increased thermal stability, high turnover number and enzymatic activity.
CC It is highly specific for the substrate and water soluble, and the
CC substrates are stable

XX
SQ Sequence 563 AA;

Query Match 99.9%; Score 2999; DB 3; Length 563;
Best Local Similarity 100.0%; Pred. No. 1.1e-236;
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVRPQRNKKRFFILINGVWNLEVTSKDRPIAVPGSWNEQYQDLCYEEGPFYKTFYVPK 60
DB 1 MVRPQRNKKRFFILINGVWNLEVTSKDRPIAVPGSWNEQYQDLCYEEGPFYKTFYVPK 60
QY 61 XLSQKHIRLYFAAVNTDCEVFLNGEKGGENHIEYLPFEVDVTGKVKSGENELRVVVENRL 120
DB 61 XLSQKHIRLYFAAVNTDCEVFLNGEKGGENHIEYLPFEVDVTGKVKSGENELRVVVENRL 120
QY 121 KVGGFPSKVPDSGTHTVGFSGFPFANPDPFPGGIIRPVLIBFTDHARILDIWDTSES 180
DB 121 KVGGFPSKVPDSGTHTVGFSGFPFANPDPFPGGIIRPVLIBFTDHARILDIWDTSES 180
QY 181 EPEKLGKVKVKEVSEAVGQEMTIKLGEEBKIRTSNRFVEGEFILLENAFWSLEDPY 240
DB 181 EPEKLGKVKVKEVSEAVGQEMTIKLGEEBKIRTSNRFVEGEFILLENAFWSLEDPY 240
QY 241 LYPKVELEKDEYTLDIGIRTISWDEKRLYLNGKPVFLKGFGKHEEPVLGGGTFFPLMI 300
DB 241 LYPKVELEKDEYTLDIGIRTISWDEKRLYLNGKPVFLKGFGKHEEPVLGGGTFFPLMI 300
QY 301 KDFNLKMINANSFRTSHYPYSEEMDLADRLGILVIDEAPHVGTIRYHNPETQKIAED 360
DB 301 KDFNLKMINANSFRTSHYPYSEEMDLADRLGILVIDEAPHVGTIRYHNPETQKIAED 360
QY 361 NIRNMIDRHKNHPSVIMSVANEPESNHPDAEGFFKALYETANENDRTRPVVMVSMMDAP 420
DB 361 NIRNMIDRHKNHPSVIMSVANEPESNHPDAEGFFKALYETANENDRTRPVVMVSMMDAP 420
QY 421 DERTRDVALKYFDIVCNRYGYWYIYQGRIBEGLOALEKDIIEELYARHKRPIFVTEFGAD 480
DB 421 DERTRDVALKYFDIVCNRYGYWYIYQGRIBEGLOALEKDIIEELYARHKRPIFVTEFGAD 480
QY 481 AIAGIHYPDPQMSBEYQAEVLEKTIKRLLLKKDYIIGTHWAFADFKTPQNVRRPILNHK 540
DB 481 AIAGIHYPDPQMSBEYQAEVLEKTIKRLLLKKDYIIGTHWAFADFKTPQNVRRPILNHK 540
QY 541 GVFTDRQPKLVAAVLRRLMSEV 563
DB 541 GVFTDRQPKLVAAVLRRLMSEV 563

RESULT 2
ADN20342
ID ADN20342 standard; protein; 563 AA.

XX AC ADN20342;
XX DT 02-DEC-2004 (first entry)
DE Bacterial polypeptide #2995.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;

KW bacterial polypeptide.

XX Bacteria.

XX US2003233675-A1.

XX 18-DEC-2003.

XX 20-FEB-2003; 2003US-00369493.

XX 21-FEB-2002; 2002US-0360039P.

XX (CAOY/) CAO Y.

XX (HINK/) HINKLE G J.

XX (SLAT/) SLATER S C.

XX (CHEN/) CHEN X.

XX (GOLD/) GOLDMAN B S.

XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX WPI; 2004-061375/06.

XX The invention relates to a recombinant DNA construct comprising a

XX promoter functional in a plant cell, where the promoter is positioned to

XX provide for expression of a polynucleotide encoding a polypeptide from a

XX microbial source. The invention also relates to a transformed plant

XX comprising the recombinant DNA construct and a method of producing a

XX transformed plant having an improved property. The plant is a crop-plant

XX such as maize or soybean. The method of producing a transformed plant

XX having an improved property comprises transforming a plant with the

XX recombinant DNA construct and growing the transformed plant, where the

XX polynucleotide or polypeptide is useful for improving plant properties.

XX The recombinant DNA construct is useful for producing plants with

XX improved plant properties, e.g. improved cold, heat or drought tolerance,

XX tolerance to herbicides, extreme osmotic conditions, pathogens or pests,

XX increased resistance to plant disease, better growth rate by modification

XX of the cell cycle pathway with plant growth regulators, increased rate of

XX homologous recombination, modified seed oil or protein yield and/or

XX content, improved yield by modification of carbohydrate, nitrogen or

XX phosphorus use and/or uptake, by modification of photosynthesis or by

XX providing improved plant growth and development under at least one stress

XX condition, improved lignin production or improved galactomannan

XX production. This sequence represents a bacterial polypeptide used in the

XX scope of the invention. Note: The sequence data for this patent did not

XX form part of the printed specification but was obtained in electronic

XX format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 563 AA;

Query Match 99.9%; Score 2999; DB 8; Length 563;
Best Local Similarity 99.8%; Pred. No. 1.1e-236;
Matches 562; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MVRPQRNKKRFFILINGVWNLEVTSKDRPIAVPGSWNEQYQDLCYEEGPFYKTFYVPK 60
DB 1 MVRPQRNKKRFFILINGVWNLEVTSKDRPIAVPGSWNEQYQDLCYEEGPFYKTFYVPK 60
QY 61 XLSQKHIRLYFAAVNTDCEVFLNGEKGGENHIEYLPFEVDVTGKVKSGENELRVVVENRL 120
DB 61 XLSQKHIRLYFAAVNTDCEVFLNGEKGGENHIEYLPFEVDVTGKVKSGENELRVVVENRL 120
QY 121 KVGGFPSKVPDSGTHTVGFSGFPFANPDPFPGGIIRPVLIBFTDHARILDIWDTSES 180
DB 121 KVGGFPSKVPDSGTHTVGFSGFPFANPDPFPGGIIRPVLIBFTDHARILDIWDTSES 180
QY 181 EPEKLGKVKVKEVSEAVGQEMTIKLGEEBKIRTSNRFVEGEFILLENAFWSLEDPY 240
DB 181 EPEKLGKVKVKEVSEAVGQEMTIKLGEEBKIRTSNRFVEGEFILLENAFWSLEDPY 240

Db 181 EPEKLGKVKIEVSEAVQEMTIKLGEEKKIRTSNRFVEGEFILENARFWSLEDPY 240

Qy 241 LYPLKVELEKDEYTLDIGRTISWDEKRLYLNGKPVFLKGFGKHSEFPVLGOGTFYPLMI 300

Db 241 LYPLKVELEKDEYTLDIGRTISWDEKRLYLNGKPVFLKGFGKHSEFPVLGOGTFYPLMI 300

Qy 301 KDFNLKWINANSFRTSHVPYSEEWLADLRILVIDEAPHVGITRHYNPETOKIAED 360

Db 301 KDFNLKWINANSFRTSHVPYSEEWLADLRILVIDEAPHVGITRHYNPETOKIAED 360

Qy 361 NRRMIDRHKNHPSVIMSVANEPESNHPDAEGFKALYETANEMDRTRPVVMVSMADAP 420

Db 361 NRRMIDRHKNHPSVIMSVANEPESNHPDAEGFKALYETANEMDRTRPVVMVSMADAP 420

Qy 421 DERTRDVALKYFDIVCVNRYGWIYQGRIEEGLQALEKDIIEELYARHRKPIFVTEFGAD 480

Db 421 DERTRDVALKYFDIVCVNRYGWIYQGRIEEGLQALEKDIIEELYARHRKPIFVTEFGAD 480

Qy 481 AIAGIHYPPOMFSEEQALVEKTIKRLKKDYIIGTHVWAFADFKTPQNVRRPILNHR 540

Db 481 AIAGIHYPPOMFSEEQALVEKTIKRLKKDYIIGTHVWAFADFKTPQNVRRPILNHR 540

Qy 541 GVFTRDQPKLVAAHLRLMSEV 563

Db 541 GVFTRDQPKLVAAHLRLMSEV 563

RESULT 3

AAW93820
ID AAW93820 standard; protein; 602 AA.

AC AAW93820;

DT 25-JUN-1999 (first entry)

DE Bacillus sp. GUS protein.

KW GUS; bogus; beta-glucuronidase; secreted; reporter molecule; marker; receptor molecule; diagnostic tool; transgene construction; plant;

KW insect; cleavage; detoxification; glucuronide.

OS Bacillus sp.

PN WO9913085-A2.

PD 18-MAR-1999.

PF 09-SEP-1998; 98WO-US019217.

PR 09-SEP-1997; 97US-0058263P.

PA (CAMB-) CAMBIA BIOSYSTEMS LLC.

PI Jefferson RA, Killian A, Keese PK;

DR WPI; 1999-229241/19.

PT New isolated microbial beta-glucuronidase.

PS Claim 4; Fig 3; 76pp; English.

CC This invention describes a novel secreted form of Bacillus sp. beta-glucuronidase (Bogus). The microbial Bogus polypeptide can be used as a reporter/effector molecule and as a diagnostic tool. The products of the invention can be used as markers for transgene constructions, e.g. in plants or insects. They can also be used for the cleavage and detoxification of glucuronides and to examine conjugation patterns of glucuronides

SO Sequence 602 AA;

Query Match 31.1%; Score 932; DB 2; Length 602;
Best Local Similarity 35.6%; Pred. No. 2.9e-67;

Matches 221; Conservative 94; Mismatches 223; Indels 82; Gaps 15;

Qy 1 MVRPQNRKKRFFILILNGVNWLEV-----TSKDRPIAVGSWNE-QYDLC 44

Db 1 MLYPINTETRGVPDLNGVWNFKLDYGKLGEEKWYBSKLTDTISMAVPSSYNDIGVTKAIR 60

Qy 45 YEEGPFYKTTFFYPKKLSQKHRLYFAAVNTDCEVFLNGEKGVENHIEYLPFEVDVTGK 104

Db 61 NHIGVYEREFVTPAYLKDQRIVLRFGSATHKAIYVNGELVVEHKGFLPFEEATINNS 120

Qy 105 VKSGENELRVVENRLKVGGFPSKVPDSGHTTVGFGS-----PPANFDFPPYG 154

Db 121 LRDMNRVTVAVDNIL-----DDSTLPVGLYSERHEEGLGKVIIRNKNPFDFFNYA 170

Qy 155 GIIRPVLIETDHAIRLIDIMWDTSESEPEKLGKVKIEVSEAVQEMTIKLG-EE 212

Db 171 GLHRPVKIYTPPTFYVEDISVTTDFNGP--TGTVYTVDFQ---GKAETVKVSVDEE 223

Qy 213 KIRTSNRFVEGEFILENARFWSLEDPYLYPLKVELEKDEYTLDI-----GRTISWDEK 267

Db 224 GKVASTEGLSGNVEIPNVILMEPLNTYLYQIKVELVNDGLTIDVVEEPFGVRTVEVNDG 283

Qy 268 RLYLNGKPVFLKGFGKHSEFPVLGOGTFYPLMIKDFNLKWINANSFRTSHVPYSEWLD 327

Db 284 KFLINNKPFYFKGFKHEDTPINGRGNFNASNVMDENILKwigANSFRTAHVPYSEELMR 343

Qy 328 LADRLGILVIDEAPHVGITRHYN-----PETOKIA-----EDNIRRMIDR 368

Db 344 LADREGILVIDETPAVGV--HLNFMAATTGLGEGSERVSTWEKIRTFEHQDVLRELVS 400

Qy 369 HKNHPSVIMSVANEPESNHPDAEGFKALYETANEMD-RTRPVVMVSMADAPDERTDV 427

Db 401 DKNHPSVVMWSIANEAAVEEGAYEYFKPLVELTKELDPQKRPTIYKFMATPE--TDK 458

Qy 428 ALKYFDIVCVNRYGWIYQGRIEEGLQALEKDIIEELYARHR----KPIFVTEFGADAIA 483

Db 459 VAEIDIVIALNRXNGWYFDGDL- --AAKVHRLQEFHAMNKRCPKIPIMITEYGADTVA 515

Qy 484 GIHYDPPOMFSEEQALVEKTIKRLKKDYIIGTHVWAFADFKTPQNVRRPILNHRGVF 543

Db 516 GFHDIDPVMFTEEYQVEYYQANHVVFDEFENFVGEQAMNFADFATSQGVMRVQGNKKGVF 575

Qy 544 TRDRQPKLVAAHLRLMSEV 563

Db 576 TRDRKPKLAAHVFRERWTNI 595

RESULT 4

AAW93825
ID AAW93825 standard; protein; 602 AA.

AC AAW93825;

DT 25-JUN-1999 (first entry)

DE Bacillus sp. codon optimised GUS protein.

KW GUS; bogus; beta-glucuronidase; secreted; reporter molecule; marker; receptor molecule; diagnostic tool; transgene construction; plant;

KW insect; cleavage; detoxification; glucuronide.

OS Bacillus sp.

PN WO9913085-A2.

PD 18-MAR-1999.

PF 09-SEP-1998; 98WO-US019217.

PR 09-SEP-1997; 97US-0058263P.

PA (CAMB-) CAMBIA BIOSYSTEMS LLC.

PI Jefferson RA, Kilian A, Keese PK;
XX WPI; 1999-229241/19.
DR N-PSDB; AAX23826.
XX
PT New isolated microbial beta-glucuronidase.
XX
PS Example 3; Fig 13A-D; 76pp; English.
XX
CC This invention describes a novel secreted form of *Bacillus* sp. beta-glucuronidase (Bogus). The microbial Bogus polypeptide can be used as a reporter/effector molecule and as a diagnostic tool. The products of the invention can be used as markers for transgene constructions, e.g. in plants or insects. They can also be used for the cleavage and detoxification of glucuronides and to examine conjugation patterns of glucuronides
CC
XX
SQ Sequence 602 AA;

Query Match 31.0%; Score 929; DB 2; Length 602;
Best Local Similarity 35.6%; Pred. No. 5.1e-67;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY 1 MVRPQRNKKRFFILINGVWNLEV-----TSKDRPIAVPGSWNE--QYQDLIC 44
DB 1 MLVPINTETRGVFDLNGVWNFKLDYGKGLEEKWYESKLTDTISMVAPSSYNDIGVTKEIR 60
QY 45 YEEGPFYKTTFFVVPKXLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTGK 104
DB 61 NHIGVWYEREFVTPAYLKDQRIVLRFSGATHKAIYVNGELVVEHKGGLPFPEAEINNS 120
QY 105 VKSGENELRVVENRLKVGFPSPKVPDSGTHVGFSGS-----FPPANFDFFPYG 154
DB 121 LRQGMNRVTVAVDNLL-----DDSTLPVGLYSERHHEGLGKVIIRNKPFDFFNYA 170
QY 155 GIIRPVLIETDHAIRLDIWDVTSESEPEKKLGKVKVIEVSEAVGQEMTIKLG--EE 212
DB 171 GLHRPVKIYTPFTYVEDISVVTDFNGP---TGTVTYTVDFQ---GKAETVKVSVVDEE 223
QY 213 KKIRTSNRFVEGEFILENARFWSLEDPLYPLKVELEKDEYTLDI----GIRTIWDEK 267
DB 224 GKVAASTEGLSGNVEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYEEFPGRVTEVNDG 283
QY 268 RLVLNGKPVFLKGFKGHEEPVLGGTFYPLMIKDFNLKMINANSFRTSHYPYSEEWLD 327
DB 284 KFLINNKPFYFKGFGKHEDTPINGRGFNEASVMDENILKWIGANSFRTAHYPSSELMR 343
QY 328 LADRLGILVIDEAPHVIGITRYHYN-----PETOKIA-----EDNIRRMIDR 368
DB 344 LADREGLVVIDETPAVG---HLNFMATTLGEGSERVSTWEKIRTFEHHQDVLRELVS 400
QY 369 HKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVMVSMMDAPDERTRDV 427
DB 401 DKNHPSVVMWSIANEATTEEGAYEYFKPLVELTKELDPQKRPTVTLVFMATPE--TDK 458
QY 428 ALKYPDIVCVNRYGWIYQGRIEEGQALEKDIIEELYARHR---KPIFVTEGADAIA 483
DB 459 VAEILDVIALNRNGWYFDGDDLAAKVLHR---QEFHAMNKRCKGKPRIMITEGADTVA 515
QY 484 GIHYDPPQMFSEEQALVEKTIKRLKKDYILGTHVWAFADFKTPQNVRRPILNHKGVF 543
DB 516 GFHDIDPVMFTEEQYVEYYQANHVFDEFENFVGEQAMNPAFATISQGMRVQGNKKGVF 575
QY 544 TRDROPKLVAVHLRLMSEV 563
DB 576 TRDRKPKLAHAVFRERWTNI 595

RESULT 5
AAW93822
ID AAW93822 standard; protein; 602 AA.
XX
AC AAW93822;

XX
DT 25-JUN-1999 (first entry)
XX
DE *Bacillus* sp. GUS protein.
XX
KW GUS; beta-glucuronidase; secreted; reporter molecule; marker;
KW receptor molecule; diagnostic tool; transgene construction; plant;
KW insect; cleavage; detoxification; glucuronide.
XX
OS *Bacillus* sp.
XX
PN WO9913085-A2.
XX
PD 18-MAR-1999.
XX
PF 09-SEP-1998; 98WO-US019217.
XX
PR 09-SEP-1997; 97US-0058263P.
XX
PA (CAMB-) CAMBIA BIOSYSTEMS LLC.
XX
PI Jefferson RA, Kilian A, Keese PK;
XX
DR WPI; 1999-229241/19.
XX
PT New isolated microbial beta-glucuronidase.
XX
PS Example 1; Fig 5; 76pp; English.
XX
CC This invention describes a novel secreted form of *Bacillus* sp. beta-glucuronidase (Bogus). The microbial Bogus polypeptide can be used as a reporter/effector molecule and as a diagnostic tool. The products of the invention can be used as markers for transgene constructions, e.g. in plants or insects. They can also be used for the cleavage and detoxification of glucuronides and to examine conjugation patterns of glucuronides
CC
XX
SQ Sequence 602 AA;

Query Match 31.0%; Score 929; DB 2; Length 602;
Best Local Similarity 35.6%; Pred. No. 5.1e-67;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY 1 MVRPQRNKKRFFILINGVWNLEV-----TSKDRPIAVPGSWNE--QYQDLIC 44
DB 1 MLVPINTETRGVFDLNGVWNFKLDYGKGLEEKWYESKLTDTISMVAPSSYNDIGVTKEIR 60
QY 45 YEEGPFYKTTFFVVPKXLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTGK 104
DB 61 NHIGVWYEREFVTPAYLKDQRIVLRFSGATHKAIYVNGELVVEHKGGLPFPEAEINNS 120
QY 105 VKSGENELRVVENRLKVGFPSPKVPDSGTHVGFSGS-----FPPANFDFFPYG 154
DB 121 LRQGMNRVTVAVDNLL-----DDSTLPVGLYSERHHEGLGKVIIRNKPFDFFNYA 170
QY 155 GIIRPVLIETDHAIRLDIWDVTSESEPEKKLGKVKVIEVSEAVGQEMTIKLG--EE 212
DB 171 GLHRPVKIYTPFTYVEDISVVTDFNGP---TGTVTYTVDFQ---GKAETVKVSVVDEE 223
QY 213 KKIRTSNRFVEGEFILENARFWSLEDPLYPLKVELEKDEYTLDI----GIRTIWDEK 267
DB 224 GKVAASTEGLSGNVEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYEEFPGRVTEVNDG 283
QY 268 RLVLNGKPVFLKGFKGHEEPVLGGTFYPLMIKDFNLKMINANSFRTSHYPYSEEWLD 327
DB 284 KFLINNKPFYFKGFGKHEDTPINGRGFNEASVMDENILKWIGANSFRTAHYPSSELMR 343
QY 328 LADRLGILVIDEAPHVIGITRYHYN-----PETOKIA-----EDNIRRMIDR 368
DB 344 LADREGLVVIDETPAVG---HLNFMATTLGEGSERVSTWEKIRTFEHHQDVLRELVS 400
QY 369 HKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVMVSMMDAPDERTRDV 427

Db 401 DKNHPSVVMWSIANEAATEEGAYEYFKPLVELTKELDPQKRPVTIVLFVMATPE--TDK 458
Qy 428 ALKTFDVCVNRYYGWYIYQGRIEGLQALEKDIIEELYARHR---KPIFVTEFGADAIA 483
Db 459 VAEILDVIALNRNGWYFDGDDLEAAKVHLR--QEFHAMNKRCPGKPIMITTEYGADTVA 515
Qy 484 GIHYDPQMFSESEYQAEIYEKTRILLKKDYIIGTHVAFADFKTPQNVRRPILNHKGVF 543
Db 516 GFHDIDPVMFTTEEYQVEYYQANHVFDEFENFVGEQAMNPFADFATSQGVMRVQGNKGVF 575
Qy 544 TRDRQPKLVAHVLRRLMSEV 563
Db 576 TRDRKPKLAHAVFRERWTNI 595

RESULT 6
AAW93826
ID AAW93826 standard; protein; 602 AA.
XX
AC AAW93826;
XX
DT 25-JUN-1999 (first entry)
XX
DE Bacillus sp. GUS protein fragment.
XX
KW GUS; BOGUS; beta-glucuronidase; secreted; reporter molecule; marker;
KW receptor molecule; diagnostic tool; transgene construction; plant;
KW insect; cleavage; detoxification; glucuronide.
XX
OS Bacillus sp.
XX
PN WO9913085-A2.
XX
PD 18-MAR-1999.
XX
PF 09-SEP-1998; 98WO-US019217.
XX
PR 09-SEP-1997; 97US-0058263P.
XX
PA (CAMB-) CAMBIA BIOSYSTEMS LLC.
XX
PI Jefferson RA, Kilian A, Keese PK;
XX
DR WPI; 1999-229241/19.
XX
PT New isolated microbial beta-glucuronidase.
XX
PS Example 3; Fig 13D; 76pp; English.
XX
CC This invention describes a novel secreted form of Bacillus sp. beta-
CC glucuronidase (BOGUS). The microbial BOGUS polypeptide can be used as a
CC reporter/effector molecule and as a diagnostic tool. The products of the
CC invention can be used as markers for transgene constructions, e.g. in
CC plants or insects. They can also be used for the cleavage and
CC detoxification of glucuronides and to examine conjugation patterns of
CC glucuronides
XX
SQ Sequence 602 AA;

Query Match 31.0%; Score 929; DB 2; Length 602;
Best Local Similarity 35.6%; Pred. No. 5.1e-67;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;
Qy 1 MVRPQNKRRFILLNGVWNLEV-----TSKDRPIAVPGSWNE--QYQDLG 44
Db 1 MLVPIINTETRGVFDLNGVWNFKLDYKGLBEKMWESKLTDTISMVAVSSYNDIGVTKEIR 60
Qy 45 YEEGPFTYKTTFYVPKLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGK 104
Db 61 NHIGVWYEREFTVPAVLKDQRIVLRFGSATHKALVYVNGELVVEHKGGLPFEEATINNS 120
Qy 105 VKSGENELRVVENRLKVGSGFSPKVPDSCGTHTVGFSGS-----FPPANFDFEPYG 154

Db 121 LRDNMRVTVAVDNIL-----DDSTLPVGLYSERHEEGLGKVI RNKNPFDFFNYA 170
Qy 155 GIIRPVLIETDHAIRLIDWDTSESEPEKLGKVKYKIEVSEAVQOEMTIKLG--EE 212
Db 171 GLHRPVKITYTTPFTVEDISVVTDFNGP--TGTVYTVDFQ---GKAETVKSVVDEE 223
Qy 213 KIRTSNRVFEGEFIIENARFWSLDDPYLPKLVELEKDEYTLDI-----GIRTSWDEK 267
Db 224 GKVASTEGLSGNVEIPNVILMEPLNTLYLQIKVELVNDGLTIDVVEEPFGVRTVEVNDG 283
Qy 268 RLYLNGKPVFLKGFGEIEFPVLGGGTFFPLMIKDFNLLKMINANSFRTSHYPYSEEWLD 327
Db 284 KFLINNKPFYFKGFGKIEDTPINGRGFNEASNVMDFNILKMI GANSFRTAHYPSSELMR 343
Qy 328 LADRLGILVIDEAPHVIGITRYHYN-----PETQKIA-----EDNIRRMIDR 368
Db 344 LADREGLVVIDETPAVG--HLNFMAATTGLGEGSERVSTWEKIRTFEHHQDVLRELVS 400
Qy 369 HKNHPSVIMWSVANEPESNHPDAEGFPKALYETANEMD-RTRPVVWVSMMDAPDERTRDY 427
Db 401 DKNHPSVVMWSIANEAATEEGAYEYFKPLVELTKELDPQKRPVTIVLFVMATPE--TDK 458
Qy 428 ALKTFDVCVNRYYGWYIYQGRIEGLQALEKDIIEELYARHR---KPIFVTEFGADAIA 483
Db 459 VAEILDVIALNRNGWYFDGDDLEAAKVHLR--QEFHAMNKRCPGKPIMITTEYGADTVA 515
Qy 484 GIHYDPQMFSESEYQAEIYEKTRILLKKDYIIGTHVAFADFKTPQNVRRPILNHKGVF 543
Db 516 GFHDIDPVMFTTEEYQVEYYQANHVFDEFENFVGEQAMNPFADFATSQGVMRVQGNKGVF 575
Qy 544 TRDRQPKLVAHVLRRLMSEV 563
Db 576 TRDRKPKLAHAVFRERWTNI 595

RESULT 7
AAB28402
ID AAB28402 standard; protein; 602 AA.
XX
AC AAB28402;
XX
DT 26-JAN-2001 (first entry)
XX
DE Staphylococcus beta-glucoronidase.
XX
KW Microbial; beta-glucoronidase; GUS; Enterobacter; Salmonella;
KW Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; bioindicator;
KW transgenic insect; marker; glucuronide detoxification.
XX
OS Staphylococcus sp.
XX
PN WO200055333-A1.
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-US007107.
XX
PR 17-MAR-1999; 99US-00270957.
XX
PA (CAMB-) CAMBIA BIOSYSTEMS LLC.
XX
PI Jefferson RA, Mayer JE;
XX
DR WPI; 2000-647075/62.
DR N-PSDB; AAA07930.
XX
PT Novel microbial beta-glucuronidase genes and gene products used as
PT reporter/effector molecule, as diagnostic tool, in positive selection, to
PT target molecules to specific cells and to detect and track linked genes.
XX
PS Example 3; Fig 3; 116pp; English.
CC The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS

D_b 360 LADREGLVIDETPAVG--HLNFMATTTGLEGSSERVSTWEKIRTFEHQDVLRELVS_r 416
Q_y 369 HKNHSVIMSVANEPBSNHPDAGFFFKALYETANEMD-RTRPVVMVSMMDAPDERTRDV 427
Db 417 DKNHSVVMWSIANEAAITEEGAYEFKPLVELTKELDPKRPTVLFWMATPE--TDK 474
Q_y 428 ALKYPDIVCVNRYYGWYIYGRIIEEGLALEKDIIEELYARHR---KPIFVTEFGADAIA 483
D_b 475 VAELIDVIALNRRYNGWYFDGGDLBAAKVHLR--QEFAHNNKRCSPGKIPMITTEYGADTVA 531
Q_y 484 GIHYDPQMFSSEYQAELVEKTIRLLKKDYIIIGHVAFADFKTPQNVRRLINHKGVF 543
D_b 532 GFHDIDPMFTEEYQVEYYQANHVFDEFENFGEQAMNPADFATSGVNRVQGNNKGVF 591
Q_y 544 TRDRQPKLVAVHLRLRMSEV 563
D_b 592 TRDRKPXLAHVFRERMTNI 611

RESULT 9
AAB28408
ID AAB28408 standard; protein; 615 AA.

AC AAB28408;

DT 26-JAN-2001 (first entry)

DE Codon-optimised *Staphylococcus* GUS.

KW Microbial; beta-glucoronidase; GUS; Enterobacter; Salmonella;

transgenic insect; marker; glucuronide detoxification.

OS Staphylococcus sp.

PN WO200055333-A1.

PD 21-SEP-2000.

PF 16-MAR-2000; 2000WO-US007107.

PR 17-MAR-1999; 99US-00270957.

PA (CAMB-) CAMBIA BIOSYSTEMS LLC.

PI Jefferson RA, Mayer JE;

DR WPI; 2000-647075/62.

XX

PT reporter/effector molecule, as diagnostic tool, in positive selection, to

[illegible]

PS Claim 3; Fig 5B; 116pp; English.

The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS genes were obtained from six different genera: *Enterobacter*/*Salmonella*, *Pseudomonas*, *Salmonella*, *Staphylococcus* and *Thermotoga*. Microbial GUS can be used as a reporter/effector molecule for transgenic constructions and in in vitro diagnostic applications. It may also be used to generate sentinel plants that serve as bioindicators of environmental status. It may be used to generate transgenic insects for tracking insect populations or to facilitate the development of a bioassay for compounds that affect molecules critical for insect development (e.g. juvenile hormone). Secreted GUS may also serve as a marker for beneficial fungi destined for release into the environment. In animal systems, secreted GUS may be used to achieve extracellular detoxification of glucuronides (e.g. toxin glucuronide) and to examine conjugation patterns of glucuronides. Microbial GUS may also be used in traditional medical diagnostic assays, for drug testing, pharmacokinetic studies, bioavailability studies, diagnosis of diseases and syndromes, following progression of disease or its response to therapy. Microbial GUS has

CC increased thermal stability, high turnover number and enzymatic activity.
CC It is highly specific for the substrate and water soluble, and the
CC substrates are stable
XX
SQ Sequence 615 AA;

Query Match	30.8%;	Score 925;	DB 3;	Length 615;
Best Local Similarity	35.7%;	Pred. No. 1.1e-66;		
Matches 220; Conservative	92;	Mismatches 223;	Indels 82;	Gaps 15;

```

QY      4 PQRNKRFFILLNGVWNLV-----TSKDRPIAVPGSWNE--QYQDLCYEE 47
      | : : ||||| :: : ||| | : : :
Db      9 PINTETRGVDFDNGVWNPFKLDYKGKLEEKYVESKLTDTTISMAVPSSYNDIGVTKEIRHI 68

```

Qy 48 GPTTYKTTFYVPKLSQKHRLYFAVNTDCEVFLNGEKVGENHIEYLPBEVDVTGKVKRS 107

Db 69 GYVMVEREFTVPAYKDKORIVLRFGSATKHAIVYVNGLVVEHKGGLFPEAEIINSLRD 128

QY 108 GENELRVVENRLKVGCFPSKVPDSCGTHTVGFCG-----FPANPDFFPYGII 157
| | : | | | : | | : | | | | :
Db 129 GMRVTVAVDNL-----DPSLT.PVGLYSRRREGI.GKVI RNKPNPDEFENYAGLH 178

QY 158 RPVLIEFTDHARILDIWDTSESEPEKKLGKVKVKIPEVSEAVGQEMTIKLG--EEEKKI 215

OY 216 RTSNRFVEGEFILENARFWSLEDPVLYPLKVELEKEYTLDI----GIRTSISWDEKRLY 270
 : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 232 VASCEGT SCANET PAUTIT WEPI NMYVOTKEU UNDCI TIDIVEEPEGDTMEVNNGKEL 291

[illegible]

Qy 331 RLGLVIDEAPHVGITRYHN-----PETOKIA----EDNIRMIDRHKH 371
||::|||||::| :|| :|::| ||
352 DEATVITDEDAVCI IIIIYMAEMOTCEGGRUJUMELVITRENEVVUDIV DEICSDP 408

QY 372 HPSTVIMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVMVSMMDAPDERTRDVALK 4300

431 YPDIVCVNRYYGWTYYQGRIEEGIALEKDIEELVARHR---KPIFYTEGADAIAGIH 486

QY 487 YDPQMFSEYQAEVETIRLLKDYIIGTHVAFADFKTPQNVRRPILNHKGVFTRD 546
| | | : | | | | : : : : | | | | | | | | | | | | | | | | | | | | | | | |

QY 547 RQPKLVAVHLRRLWSEV 563
|:| | | | | : :
|:| | | | | : :
|:| | | | | : :

RESULT 10

ID AAB28409 standard; protein; 602 AA.

AC AAB28409;

DT 26-JAN-2001 (first entry)

Salmonella beta-glucuronidase.

KW Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;

transgenic insect; marker; glucuronide detoxification.

OS *Salmonella* sp.

PN WO200055333-A1.

PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-US007107.
XX
PR 17-MAR-1999; 99US-00270957.
XX
PA (CAMB-) CAMBIA BIOSYSTEMS LLC.
XX
PI Jefferson RA, Mayer JE;
XX
DR WPI: 2000-647075/62.
DR N-PSDB; AAA07939.
XX
PT Novel microbial beta-glucuronidase genes and gene products used as
PT reporter/effector molecule, as diagnostic tool, in positive selection, to
PT target molecules to specific cells and to detect and track linked genes.
XX
PS Example 3; Fig 17; 116pp; English.
XX
CC The present sequence is a microbial beta-glucuronidase (GUS) protein. GUS
CC genes were obtained from six different genera: Enterobacter/Salmonella,
CC Pseudomonas, Salmonella, Staphylococcus and Thermotoga. Microbial GUS can
CC be used as a reporter/effector molecule for transgenic constructions and
CC in vitro diagnostic applications. It may also be used to generate
CC sentinel plants that serve as bioindicators of environmental status. It
CC may be used to generate transgenic insects for tracking insect
CC populations or to facilitate the development of a bioassay for compounds
CC that affect molecules critical for insect development (e.g. juvenile
CC hormone). Secreted GUS may also serve as a marker for beneficial fungi
CC destined for release into the environment. In animal systems, secreted
CC GUS may be used to achieve extracellular detoxification of glucuronides
CC (e.g. toxin glucuronide) and to examine conjugation patterns of
CC glucuronides. Microbial GUS may also be used in traditional medical
CC diagnostic assays, for drug testing, pharmacokinetic studies,
CC bioavailability studies, diagnosis of diseases and syndromes, following
CC progression of disease or its response to therapy. Microbial GUS has
CC increased thermal stability, high turnover number and enzymatic activity.
CC It is highly specific for the substrate and water soluble, and the
CC substrates are stable
XX
SQ Sequence 602 AA;

Query Match 30.7%; Score 920; DB 3; Length 602;
Best Local Similarity 35.3%; Pred. No. 2.8e-66;
Matches 213; Conservative 110; Mismatches 223; Indels 58; Gaps 11;
QY 1 MVRPQRNKKRFILINGVWNLVETSKD-----RPIAVPGSWNEQY--QDL 43
DB 1 MLRSVETATREIKKLDGLMSFCMDSBECGNAQQWROPLPQSRALAVPGSYNDQRAAEI 60
QY 44 CYEEGPFYKTFYVPKLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPEFVDVTG 103
DB 61 RNYGVNVWYQREIRIPKGDWRQRIYLRFDVATHYGVWVNDQFLMEHQGGYTPFADISH 120
QY 104 KVKSGEN-ELRVVVENRLKVGFPKVPDSGTHVGFPGSPPPANF-DFPPYGGIIRPVL 161
DB 121 LISAGESVRITVCVNNELMWQTI-----PGVVTGGVNGKKQQAAYFHDFFNYAGIHR SVM 175
QY 162 IEFTHARILDIWVDTSESEPEKKLGKVKKIEVSEEA VGQEMTIKLGEBEKKIIRTSNR 221
DB 176 LYTPPKTFVEDITVTQVAD--DLAQATVAMQVRANG--EVRVELRDAEQQLVASGQG 229
QY 222 VEGEFILENARFWSLEDPYLYPLKV----ELEKDEYTLDIGIRTSWDEKRLYLNGKPV 277
DB 230 EKGELLLEGRLWQPEGYLYELRYLAQHODEQDEYPLRVGIRSEVKGEOFLINH KPFY 289
QY 278 LKGFGRHEEPVVLGGTFYPLMIKIDENLLKMINANSFRTSHYPYSEEWL DLAIRLGI LVI 337
DB 290 FTGFRGHEDADLRGKGFNDVLMVHDALMDWIGANSYRTSHYPYAEMLDWADEHGI VTI 349
QY 338 DEAPHVGI-----TRYHNPETQKIAEDNIRRMIDRHKNHPSVIMWS 379
DB 350 DETAAVGFNLGLISFDVGEKPKELYSDA VNDETQRAHLQAIKELIARDKNHPSVIMWS 409

QY 380 VANEPESNHPDAEGFFKALYETANEMDRTPVWVSM--DAPDERTRDVALKFYDVCV 437
DB 410 IANEPDRPNGAREYFAPLAQTRELDPTRPTCVNMFCDAESDITIDL----FDV VCL 465
QY 438 NRYYGWYIYQGRIEGLQALEKDIEELYARHRKPIFVTEFGADALAGIHYPQMFSEEX 497
DB 466 NRYYGWYVQSGDLEKAEKVELKELLAWQEKLRPIIITEYGVDTLAGLSMVDMSSEEX 525
QY 498 QAEVLEKTIIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHKGVFTRDRQPLVAHVLR 557
DB 526 QCAMLDMYHRVFDRVSAVVGEOVWNFADFATSGIMRVGNNKGIFTRDRKPKSA AFLQ 585
QY 558 RLWS 561
DB 586 KRW 589

RESULT 11
AAM48998
ID AAM48998 standard; protein; 598 AA.
XX
AC AAM48998;
XX
DT 10-MAY-2002 (first entry)
XX
DE L gasserii beta-glucuronidase.
XX
KW Beta-glucuronidase; GUS; low pH; Lactobacillus.
XX
OS Lactobacillus gasserii.
XX
PN WO200190305-A2.
XX
PD 29-NOV-2001.
XX
PF 22-MAY-2001; 2001WO-US016667.
XX
PR 23-MAY-2000; 2000US-0206372P.
XX
PA (UYNC-) UNIV NORTH CAROLINA STATE.
XX
PI Russell WM, Klaenhammer TR;
XX
DR WPI: 2002-062529/08.
DR N-PSDB; AAL44836.
XX
PT Novel isolated nucleic acid of Lactobacillus gasserii encoding beta
PT glucuronidase having activity at acid pH which is useful as reporter
PT protein for highly aciduric organisms, or as marker of transformed cells.
XX
PS Claim 19; Page 49-51; 53pp; English.
XX
CC The present invention provides the protein and coding sequences of
CC Lactobacillus gasserii beta-glucuronidase (GUS). The protein has maximum
CC activity at acidic pHs, and thus can be used to detect low pH
CC environments. The present sequence is the protein of the invention
XX
SQ Sequence 598 AA;
Query Match 30.6%; Score 919.5; DB 5; Length 598;
Best Local Similarity 34.1%; Pred. No. 3e-66;
Matches 207; Conservative 110; Mismatches 227; Indels 63; Gaps 14;
QY 4 PQRNKKRFILINGVWNLVET-----SKDRP----I AVPGSWNE--QYQDL CYEE 47
DB 7 PIQNKYRFNTLMNGTWQFETDPSVGLDEGWNKELPDPEEMVVPGTFAELTTKRDRKYTT 66
QY 48 GPFTYKTFYVPKLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPEFVDVTGK VKS 107
DB 67 GDFWYQKDFIIPSLKKKELYIRFGSVTHRAKVFINGHEVGQHEGGFLPFQV KISNYIN Y 126
QY 108 GE-NELRVVVENRLKVGFP---SKVPDSGTHTVGFSGSPPPANFDFPPYGGIIRPVLIE 163

Db 127 DQTRVTVLVNNELSEKALPCGTEILDNGQ-----KLAQPYDFEFNYSGIMRNWLL 179
Qy 164 FTDHARLDIWDVDTSESEPEKLGKVKYKIEVSEAVGQ-EMTIKLGEEKKIRTSNRFV 222
Db 180 ALPOSQITNFKLN-----YQLANNKATITYNIEANNNAEFKVTLPFQNGKEVACATSKN 232
Qy 223 EGEFILENARFWSLDEPPLYPLKVELEK-----DEYTLDIGIRTISWDEKRLYLNGKVF 277
Db 233 TSSLITGNPHLMSPNDRPYKIKIEMLEDEGKTVDYETDKIGIRTVKIKINDKILNNHPIY 292
Qy 278 LKGFQKHKEFPVLGGCTFYPMLIKDFNLKWINANSFRTSHYPSSEWLDLADRLGLVI 337
Db 293 LKGFQKHEDFNLGKAVNESIIRKDYECMKKIGANCRRSSHYPYAEWYQYADKYGFLII 352
Qy 338 DEAPHVGITRYHYN-----PETOKIAEDNIRRMIDRHKNHPSYIMWS 379
Db 353 DEVPAVGLNRSITNPLVNTNSQSHFPAASKTVPELKVHEQEIKEWIDRDQRHPSVIAWS 412
Qy 380 VANEPESNHPDAEGFFKALYETANEMD-RTRPVVMVSMMDAPDERTRDVALKYFDIYCVN 438
Db 413 LFNEPESTTQESYDYFKDIFAFARKLDQNRPTYGTILVMSGPK--VDKLPCLDFVCLN 470
Qy 439 RYRGWYIYQG-RIEEGLALEKDIIEELY-ARHRKPIFVTEFGADALAGIHYDPQPMSEE 496
Db 471 RYRGWYVAGGPEIVNAKKMLEDELQWONLKLKPFVFTFEGADTLSSSHRLPDEWMSQE 530
Qy 497 YQAEIVKTIIRLLKKDYIIGTHVWAFADFKTPQNVRRPILNHKGVFTRDQPKLVAVHL 556
Db 531 YQNEYQMYFDIFKKYFPICGELVWNPADFKTSEGIMRVGCGNDKGIPTRDREPKDIAFTL 590
Qy 557 RRLWSEV 563
Db 591 KKRWQQL 597

RESULT 12
ADQ89612
ID ADQ89612 standard; protein; 656 AA.
XX AC ADQ89612;
XX DT 21-OCT-2004 (first entry)
XX DE Antagonist of cell cycle progression polypeptide #21.
XX KM Cytostatic; cancer; cell division cycle; mitosis; meiosis;
XX KM cell cycle progression.
XX OS Drosophila melanogaster.
XX PN WO2004063362-A2.
XX PD 29-JUL-2004.
XX PF 31-DEC-2003; 2003WO-GB005635.
XX PR 10-JAN-2003; 2003US-0439123P.
XX PR 06-MAY-2003; 2003US-0468402P.
XX PA (CYCL-) CYCLACEL LTD.
XX PI Glover D, Bell G, Frenz L, Midgley C;
XX DR WPI: 2004-544089/52.
XX DR N-PSDB; ADQ89611.
XX PT New cell cycle progression genes and proteins for modulating cell cycle
XX PT progression in cells, for preventing, treating or diagnosing cell
XX PT proliferative diseases (e.g. cancer) or for identifying modulators of
XX PT mitosis or meiosis.
PS Claim 2; SEQ ID NO 42; 461bp; English.

XX The present invention relates to a polynucleotide for preventing,
CC treating or diagnosing a disease in an individual. The composition or the
CC polypeptide, polynucleotide or RNA precursor, or antibody is useful for
CC diagnosing, preventing or treating diseases (e.g. cell proliferative
CC diseases such as cancer) in an individual. These may also be used for
CC identifying substances capable of binding to or modulating the function
CC of the polypeptide, capable of affecting the function of the
CC corresponding gene, or capable of inhibiting the cell division cycle or
CC cell cycle progression, preferably mitosis and/or meiosis. The present
CC sequence represents an antagonist of cell cycle progression protein
CC
XX
SQ Sequence 656 AA;

Query Match 30.2%; Score 905; DB 8; Length 656;
Best Local Similarity 34.7%; Pred. No. 5.3e-65;
Matches 212; Conservative 122; Mismatches 209; Indels 68; Gaps 17;

Qy 1 MVRPQRNKKRFLILNGVNL-----EVTSKDR---PIAVPGSWNEQ 39
Db 34 MLYRESETRVRSIDGIMNPFVRSDQANPTQGVDEWYAKELSKSRPTIMPVPASYNDI 93
Qy 40 YQD-LCYEEGPFYKTTFFYPVKLSQ-KHIRLYPAAVNTDCEVFLNGEKVGENHIEYLPF 97
Db 94 TTDLNRDHDVGTWYDRKFFVPRFSKQDRIMWLRFGSVHYEAYVWINGQKVKGHEGLHP 153
Qy 98 EVDVTGKVKSG-ENEELRVVENRLKVGCFP---SKVPDSGTHTVGFFGSFPFANDFPP 152
Db 154 EAEVTDLISYGAENRITVMCDNALIQTTVPQGRITEVPNDGWTI----VQSYTFDFFN 208
Qy 153 YGGIIRPVLIIEFTHARILDIWDVDTSESEPEKLGKVKYKIEVSEAVGQ----EMTIK 207
Db 209 YAGIHRSVHLTYTPRTFIEEVEVTNLKSK-DATVGEVYVSVNGSAANEADNVLQIQAN 267
Qy 208 LGEEBKII--RTSNRFVEGEFILENARFW-----SLEDPLYPLKVELEK-----DEY 253
Db 268 LYDKDGLVANATSDQKLGKLYQNPVKPWPVYLMHSEPGYLYQIEIKLATNDELDLVY 327
Qy 254 TLDIGIRTISWDEKRLYLNGKPVFLKGFQKHKEFPVLGGTFYPLMIKDFNLKWINANS 313
Db 328 RLKVGIRTLSSWSQQLFINGKPYFRGFRHEDSDIRGKGLDNALMVRDFNLKWI GANA 387
Qy 314 FRTSHYPSSEWLDLADRLGILVIDEAPHVGITRYHYNPETOKIAEDNIRRMIDRKNHP 373
Db 388 YRTSHYPSSESMQFADEHGIMIDECPVDTE--NFSQELGKHKSSLEQLIHRDRNP 445
Qy 374 SVIMSVANEPESNHPDAEGFFKALYETANEMDRTRPVVMVSMMDAPDERTRDVALKYFD 433
Db 446 SVVWMSIANEPRGTSVSADSYFELVANFTRSLDKTRPITAIAV----SNTQDKAGRSLD 501
Qy 434 IVCVNRYYGWYIYQGRIEGLQALEKDIIEELYA---RHRKPIFVTEFGADALAGIHYDP 490
Db 502 IISFNRYNAMYSNAGRLD--MITQNVIDEALMNKRYNKPIMSEYGADTLEGLHMQPA 558
Qy 491 QMESEYQAEIVKTIIRL---LKKDYIIGTHVWAFADFKTPQNVRRPILNHKGVFTRDR 547
Db 559 YVWSEEFQTEVFSHPKAFDELRRKGWFIGEFVWNPADFXTAGSYTRVGKNGKGVFTRAR 618
Qy 548 QPKLVAAHLRR 558
Db 619 QPKAAHLRLRK 629

RESULT 13
ADF28924
ID ADF28924 standard; protein; 648 AA.
XX ADF28924;
XX AC ADF28924;
XX DT 12-FEB-2004 (first entry)
XX DE Murine beta-glucuronidase precursor.

QY 414 VSMMDAPDERTRDVALKYFDIVCVNRYYGWYIYQGRIEEGLQALEKDIIELYA---RHRK 470
Db 473 AIAV-----SNTQDKAGRSLDIISFNRYNAMYSNAGRLD--MITONVIDEAIAMNKRYNK 525
QY 471 PIFVTEFGADAIAGIHYDPQPMFSEYQAEVVEKTI RL--LLKKDYIIIGTHVWAFADFK 527
Db 526 PIMSEYGADTLEGHMQPAYVWSEEFQTEVFSRHPKAFDELKKGWFIGEFVWNPADFK 585
QY 528 TPQNVRRPILNHNKGVFTRDRQPKLVAVHLRR 558
Db 586 TAQSYTRVGGNKKGVFTRARQPKAAAHLLRK 616

Search completed: September 1, 2005, 19:29:37
Job time : 172 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2005, 19:22:34 ; Search time 42 Seconds
(without alignments)
1289.763 Million cell updates/sec

Title: US-09-936-759-6

Perfect score: 3001
Sequence: 1 MWRPQNKRRFILNGVN.....TRDRQKLVAVHLRLMSEV 563

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 bsgs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

```

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

```

```

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```
Database : PIR_79: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2999	99.9	563	2	A72300	beta-glucuronidase
2	1011	33.7	570	2	C90485	beta-glucuronidase
3	899.5	30.0	648	2	A32576	beta-glucuronidase
4	894.5	29.8	648	2	A25047	beta-glucuronidase
5	879	29.3	603	1	GBECCG	beta-glucuronidase
6	865.5	28.8	651	2	A26581	beta-glucuronidase
7	691	23.0	368	2	A85768	partial beta-D-glu
8	691	23.0	370	2	D90919	beta-D-glucuronida
9	487	16.2	755	2	D95842	probable beta-gala
10	482	16.1	716	2	JU0275	beta-galactosidase
11	481.5	16.0	996	2	D86872	beta-galactosidase
12	452.5	15.1	1087	2	F72283	beta-galactosidase
13	414	13.8	1024	1	GBEC	beta-galactosidase
14	405.5	13.5	1024	2	E90678	beta-D-galactosida
15	405.5	13.5	1024	2	A85529	beta-D-galactosida
16	381.5	12.7	1014	2	C83990	beta-galactosidase
17	371.5	12.4	1307	2	T35944	probable beta-gala
18	364	12.1	897	2	A39405	beta-galactosidase
19	360.5	12.0	1060	2	A10201	beta-galactosidase
20	352.5	11.7	1025	1	JC1266	beta-galactosidase
21	352.5	11.7	1042	1	GBECE	beta-galactosidase
22	352.5	11.7	1042	2	E85968	evolved beta-D-gal
23	352.5	11.7	1042	2	F91123	beta-galactosidase
24	348.5	11.6	1034	2	T30574	beta-galactosidase
25	348	11.6	1026	2	A49750	beta-galactosidase
26	339.5	11.3	1034	2	A24925	beta-galactosidase
27	328	10.9	1035	2	T47603	beta-galactosidase
28	327.5	10.9	1034	2	T30551	beta-galactosidase
29	322.5	10.7	626	2	A42891	beta-galactosidase

30	321	10.7	2228	2	E97942	beta-galactosidase
31	321	10.7	2233	2	B95075	beta-galactosidase
32	312	10.4	1015	2	I39697	beta-galactosidase
33	310	10.3	1005	2	T31333	beta-galactosidase
34	300.5	10.0	1007	2	A30093	beta-galactosidase
35	256	8.5	785	2	H72228	hypothetical prote
36	194	6.5	237	2	B90919	interrupted beta-D
37	194	6.5	237	2	B85768	partial beta-D-glu
38	191.5	6.4	827	2	AB2764	beta-mannosidase p
39	191.5	6.4	832	2	A97540	hypothetical prote
40	181	6.0	820	2	T37230	probable beta-mann
41	179.5	6.0	900	2	T19689	hypothetical prote
42	179	6.0	879	2	A55881	beta-mannosidase (
43	175	5.8	891	2	A82755	beta-mannosidase p
44	139.5	4.6	1091	2	F83928	hypothetical prote
45	132	4.4	818	2	A97668	mannosidase AGR C

ALIGNMENTS

RESULT 1

beta-glucuronidase - *Thermotoga maritima* (strain MSB8)

C;Species: Thermotoga maritima

C;Date: 11-Jun-1999 #sequence revision 11-Jun-1999 #text change 09-Jul-2004

C;Accession: A72300

R; Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, R.; Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.C.M.

Nature 399, 323-329, 1995

A,Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing
A,Reference number: A72200; MUID:99287316; PMID:10360571

A;Accession: A72300

A;Status: preliminary

A:Molecule type: DNA

A;Residues: 1-563 <ARN>

A; Cross-references: UNI

A; Experimental source: strain MSB8

C;Genetics:

A;Gene: TM1062

C;Superfamily:

Order Match

Query Match	99.9%;	Score 2999;	DB 2;	Length 563;
Best Local Similarity	99.8%;	Pred. No. 4,2e-197;		
Matches 562;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0
QY 1	MVRPQRNKKRPFILILINGVMNIEVTSKDRPIAVPGSWNEQYODICYEEGPFTYKTTFFVVK	60		
DB 1	MVRPQRNKKRPFILILINGVMNIEVTSKDRPIAVPGSWNEQYODICYEEGPFTYKTTFFVVK	60		
QY 61	XLSEQHIRLYFAAVNTDCEVFINGEKVGENHIEYLPFEVDVTGKVKSGENELRVVENRL	120		
DB 61	ELSEQHIRLYFAAVNTDCEVFINGEKVGENHIEYLPFEVDVTGKVKSGENELRVVENRL	120		
QY 121	KVGGFPSKVPDPSGTHTVGFPGSPFPANFDFPYGCIIRPVLIEFTDHARILDIWDTSES	180		
DB 121	KVGGFPSKVPDPSGTHTVGFPGSPFPANFDFPYGCIIRPVLIEFTDHARILDIWDTSES	180		
QY 181	EPEKLGKVKVKEI EVSEEA VQGE MTKLGEEKKIRTSNRFVEGEFIL ENARFMSLED PY	240		
DB 181	EPEKLGKVKVKEI EVSEEA VQGE MTKLGEEKKIRTSNRFVEGEFIL ENARFMSLED PY	240		
QY 241	LYPLKVELKDEYTLDIGIRTISWDEKRLYLNGKPVFLKGFGRHEEFVPLGGGTFFPLMI	300		
DB 241	LYPLKVELKDEYTLDIGIRTISWDEKRLYLNGKPVFLKGFGRHEEFVPLGGGTFFPLMI	300		
QY 301	KDFNLKWINANSFRTSHYPYSEEWLDLADRLGILVIDEAPHVGITRHYHNPETOKIAED	360		
DB 301	KDFNLKWINANSFRTSHYPYSEEWLDLADRLGILVIDEAPHVGITRHYHNPETOKIAED	360		
QY 361	NIRRMIDRHKNHPSVIMMSVANE PESNH PDAGCFKALYETANEMDRTRPVVAVWSMDAP	420		

Db 361 NRRMIDRHKHNP SVIMSVANEPESNHPDAEGFFKALYETANEMDRTRPVVMVSMMDAP 420

QY 421 DERTRDVALKFDIVCVNRYGWIYQGRIEBGLQALEKDI EELYARHKRPIFVTEFGAD 480

Db 421 DERTRDVALKFDIVCVNRYGWIYQGRIEBGLQALEKDI EELYARHKRPIFVTEFGAD 480

QY 481 AIAGIHYDPQPMFSEEQALVEKTRILLKKDYIIGTHVWAFADFKTPQNVRRPILNHK 540

Db 481 AIAGIHYDPQPMFSEEQALVEKTRILLKKDYIIGTHVWAFADFKTPQNVRRPILNHK 540

QY 541 GVFTDRQPKLVAAHVLRLMSEV 563

Db 541 GVFTDRQPKLVAAHVLRLMSEV 563

RESULT 2

C90485
beta-glucuronidase (gusb) [imported] - Sulfolobus solfataricus
C/Species: Sulfolobus solfataricus
C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C/Accession: C90485
R.;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F
arrett, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.
submitted to Genbank, April 2001
A/Description: Sulfolobus solfataricus complete genome.
A/Reference number: A99139
A/Accession: C90485
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-570 <KUR>
A/Cross-references: UNIPROT:Q97U11; GB:AB006641; NID:g13816434; PIDN:AAK43138.1; GSPDB:G
C/Genetics:
A/Gene: gusb
C/Superfamily: beta-glucuronidase

Query Match 33.7%; Score 1011; DB 2; Length 570;
Best Local Similarity 39.2%; Pred. No. 2.7e-61;
Matches 230; Conservative 92; Mismatches 191; Indels 74; Gaps 14;

QY 15 LNVNMLEVTSKDRP-----IAPGSWNEQYQDLCYEEGPFYKTFYVPK 60

Db 11 LQGFWKFKIDNENTGENGWYKGLSEDIIVPASWNEQNPKWDFSGIAWYQKDLFVSN 70

QY 61 XLSQHIRLYFAAVNTDCEVFLNGEKYGENHIEYLPDEVDTGKYKSGENLRVVVENRL 120

Db 71 DNGNRKAMWVFEGAGYITKLWINGEGCTHESFTQPKPIKLKV----NEFNKIV--V 123

QY 121 KVGFPESKVPDSGITHVGFFGSFPPAN-----FDFPPYGGIIRPVLIETDHARILDIW 174

Db 124 KIDNTFSPY-----NLPPARDLNNAAFDFPNYGGIHRPYIIEFVDECHVEDIT 171

QY 175 VDTSESEPEKLGKVKYKIEVSEAVQEMTIKLGESEKKI---RTSNRFVEGEFLEN 230

Db 172 VYT-----KSYGHLKVEI-LSECNQRFSLRFKLVKGRVILNESSNEVFEDK--VNN 222

QY 231 ARFWSLEDPPYLYPLKVELE----KDEYTLDIGIRTSWDEKRLYLNGKPVFLKFGKGHE 285

Db 223 VIPWSPDNPYLYTLIVEMYVVGNLKDSYVERIGFRDVEVKDGKIYLNKGPFIFLKFGGRHE 282

QY 286 EFPVLQCGTFYPLMIKDNLLKWINANSFRTSHYPYSEEWLADRLGILVIDEAP--HV 343

Db 283 DFPILGKFTYGAVLVDFYLMRKIGANSFRTSHYPYSENEHLADLDEMGFLVILEPPLCY 342

QY 344 GITRHYNPETOKI-----AEDNIRRMIDRHKHNP SVIMSVANEPESNHPDAEGF 394

Db 343 NISRWMSQEI AKMFGDVKYFEKVRDTIKEMIRQHKNRPSVIMYSVWNEPPSDIREVAEF 402

QY 395 FKALYETANEMDRTRPVVMVSMMDAPERTRDVALKFDIVCVNRYGWIYQGRIEGL 454

Db 403 IRREVELFKSLDSSRPYTFAS-----HRSVRDLALEYDVISLNYHGMWYTEMWGDI DSGV 457

QY 455 QALEKDI EELYARH-RKPIFVTEFGADAIAGIHYDPQPMFSEEQALVEKTRILLKKD 513

Db 458 KVAIAIELEIHKFPEKPIITTEFGADAIYGLHSDPFQMWSEEQSEMIRKYIEALREKD 517

QY 514 YIIGTHVWAFADFKTPQNVRRPILNHKGVFTDRQPKLVAAHVLRLM 560

Db 518 YIVGFHIWNAFDRTPQNPSPRITLNRKGIPTDRQPKLAAKVEELF 564

RESULT 3

A32576
beta-glucuronidase (EC 3.2.1.31) allele B precursor - mouse
N/Alternate names: beta-D-glucuronoside glucuronosohydrolase
C/Species: Mus musculus (house mouse)
C/Date: 12-Oct-1989 #sequence_revision 12-Oct-1989 #text_change 09-Jul-2004
C/Accession: A32576; B32576; I49692; A28954; A29977; A35798
R.;Wawrzyniak, C.J.; Gallagher, P.M.; D'Amore, M.A.; Carter, J.B.; Lund, S.D.; Rinchik, B
Mol. Cell. Biol. 9, 4074-4078, 1989
A/Title: DNA determinants of structural and regulatory variation within the murine beta-
A/Reference number: A32576; MUID:89384641; PMID:2779578
A/Accession: A32576
A/Molecule type: mRNA
A/Residues: 1-648 <WAW>
A/Cross-references: UNIPROT:P12265; GB:M28540; GB:M27816; NID:g193718; PIDN:AAA63307.1;
A/Experimental source: allele B
A/Accession: B32576
A/Molecule type: mRNA
A/Residues: 1-86,'I',88-648 <WA2>
A/Cross-references: GB:M28541; NID:g193720; PIDN:AAA63308.1; PID:g193721; GB:M27816
A/Experimental source: allele H
R.;Funkenstein, B.; Leary, S.L.; Stein, J.C.; Catterall, J.F.
Mol. Cell. Biol. 8, 1160-1168, 1988
A/Title: Genomic organization and sequence of the Gus-s-a allele of the murine beta-gluc
A/Reference number: I49692; MUID:88216590; PMID:2835664
A/Accession: I49692
A/Status: translated from GB/EMBL/DDBJ
A/Molecule type: mRNA
A/Residues: 1-232,'T',234-264,'D',266-319,'V',321-427,'K',429-615,'L',617-648 <RES>
A/Cross-references: GB:M19279; NID:g193524; PIDN:AAA37697.1; PID:g309257
A/Experimental source: allele A
R.;D'Amore, M.A.; Gallagher, P.M.; Korfhagen, T.R.; Ganschow, R.E.
Biochemistry 27, 7131-7140, 1988
A/Title: Complete sequence and organization of the murine beta-glucuronidase gene.
A/Reference number: A28954; MUID:89062453; PMID:3196706
A/Accession: A28954
A/Molecule type: DNA
A/Residues: 1-264,'D',266-319,'V',321-648 <DAM>
A/Cross-references: GB:J02836; GB:J03035; GB:M20204; NID:g193716; PIDN:AAA98623.1; PID:g
R.;Gallagher, P.M.; D'Amore, M.A.; Lund, S.D.; Ganschow, R.E.
Genomics 2, 215-219, 1988
A/Title: The complete nucleotide sequence of murine beta-glucuronidase mRNA and its dedu
A/Reference number: A29977; MUID:88284700; PMID:3397060
A/Accession: A29977
A/Molecule type: mRNA
A/Residues: 1-264,'D',266-319,'V',321-648 <GAL>
A/Cross-references: GB:J03047; NID:g193522; PIDN:AAA37696.1; PID:g309256
R.;Li, H.; Takeuchi, K.H.; Manly, K.; Chapman, V.; Swank, R.T.
J. Biol. Chem. 265, 14732-14735, 1990
A/Title: The propeptide of beta-glucuronidase. Further evidence of its involvement in co
the serpin superfamily.
A/Reference number: A35798; MUID:90368633; PMID:2394691
A/Accession: A35798
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 593-648 <LIA>
A/Note: the location of the propeptide cleavage site was not demonstrated directly but r
C/Comment: In some tissues, a portion of this enzyme is retained in the endoplasmic reti
syn.
C/Genetics:
A/Gene: Gus
A/Map position: 5
A/Intons: 70/3; 132/3; 193/2; 241/1; 303/3; 351/3; 411/2; 460/2; 488/3; 547/3; 593/1
C/Superfamily: beta-glucuronidase
C/Keywords: glycosidase; hydrolase; lysosome

Db	147	LPFEADISNLVQVGLPSPRLRITIAINNLT-----PTLPBGTIQYLTDTSKYPKGYF--	200
Qy	143	FPFANFDFEPPYGGIIRPVLIETDARIIDIMVDTSESEPEKKLGKVKVKEIVSEAVGQ	202
Db	201	VQNTYFDFFNVAAGLQRSVLLYTTPTTYIDITVTS---VEQDSGLVNVQISVKSNNL-F	256
Qy	203	EMTIKLGSEEEKIRTSNRFVEGEFILENARFW---SLEDP-VLYPLKYLE-----	249
Db	257	KLEVRLLDAENKVVANGTQGGQLKPGVSLMWPYLMHERPAVLYSLEVQLTAQTSIGPV	316
Qy	250	KDEYTLDIGIRTISWDEKRLYLNGKPVFLKFGKHEEFVVGQTFYPLMIKDFNLKMI	309
Db	317	SDFYTLPGVIRTVAVTKSQFLNGKPFYFHVGNKHEDADIRGKGFDPVLLVKDFNLRL	376
Qy	310	NANSFRSHYPPYSEEWLADRLGILVIDEAPHV-----ITRYHNPETOKIAED	360
Db	377	GANAFRSHYPPYAEVWQMDRXYIVIDECPVGLALPQFPNNVSLHH---MQVME	432
Qy	361	NIRRMIDRHKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMDRTRPVVMVSMADP	420
Db	433	VVR----RDKNHPAVVMVSVANEPASHLESAGYLLKMYAHTKSLDSPSPVTFVS---N	484
Qy	421	DETRDVALKFDIVCVNRYGWIYQGRIEEGLQALEKDIEELYARHKRPIFTEFGAD	480
Db	485	SNYAADKGAAPYVDVICLANSYSWYHDYGHLELIQLQLATQFENMYKKYQKPIIQSEYGA	544
Qy	481	AIAGIHNDPQPMFSEYQALVEKTIRLL--KDYIIGTHVWAFADKTPQNVRRPIL	537
Db	545	TIAGFHQDPPLMFTEEYQXSLLEQ-YHLGLDQKRKYVVGELIWNFADPMTQSPTRVLG	603
Qy	538	NHKGVFTDRQPKLVANVLR-RLW	560
Db	604	NKKGIFTRQRPKSAAPLLRERYW	627

RESULT 7
A85768
partial beta-D-glucuronidase [imported] - Escherichia coli (strain O157:H7, substrain ED
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: A85768
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Hiller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85768
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-368 <STO>
A:Cross-references: UNIPROT:Q8X671; GB:AE005174; MID:G12515602; PIDN:AA656605.1; GSPDB:C
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: uidA_1

Query Match	23.0%;	Score 691;	DB 2;	Length 368;
Best Local Similarity	39.6%;	Pred. No. 1e-39;		
Matches	141;	Conservative	62;	Mismatches 125; Indels 28; Gaps 4;
Qy	230	NARFWSLEDPYLYPL----KVELEKDEYTLDIGIRTISWDEKRLYLNGKPVFLKFGKHE	285	
Db	3	NPHLMQGEGLYELLYTAKSRTCEDIYPLRVGIRSAVAKGEQFLNHKPFYTFGFGHE	62	
Qy	286	EFPVLGQGTFFYPLMIKDFNLKWINANSFRTSHYPYSEEWLADRLGILVIDEAPHVGI	345	
Db	63	DADLRGKGFNDVLMVHDHALMDWIGANSYRTSHYPYAEEMLDWADEHGIVIDETAAGVF	122	
Qy	346	-----TRYHNPETOKIAEDNIRRMIDRHKNHPSVIMSVANEPESN	387	
Db	123	NLSLIGTFEAGNKPKEYSEEA VNGETQQAHLQAIKELIARDKNHPSVVMMSIANEPDTR	182	
Qy	388	HPDAEGFFKALYETANEMDRTRPVVMVSM--DAPDETRDVALKYPDIVCVNRYGWI	445	

Db	183	PGGAREYFAPLAEATRKLDPTRPITCVNMFCDAHTDTISDL-----FDVLCNRRYGYWYV	238
Qy	446	YQGRIEEGLQALEKDIEELYARHKRPIFVTEFGADAIAGIHYPDPQMFSEYQALVEKT	505
Db	239	QSGDLTAEKVLLEKELLAWQEKLHQPIITTEYGVDTLAGLHSMYTDMMSEEYQCAWDMY	298
Qy	506	IRLLKDDYIIGTHVWAFADKTPQNVRRPILNHKGVFTDRQPKLVANVLRRLWS	561
Db	299	HRVFDRAVSAVGEQVWNPADFATISQILRVGKNKGIFTRDRKPKSAAPLLQKRWT	354

RESULT 8
D90919
beta-D-glucuronidase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05091
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: D90919
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D90919
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-370 <HAY>
A:Cross-references: UNIPROT:Q8X671; GB:BA000007; PIDN:BA635747.1; PID:G13361791; GSPDB:C
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs2324

Query Match	23.0%;	Score 691;	DB 2;	Length 370;
Best Local Similarity	39.6%;	Pred. No. 1e-39;		
Matches	141;	Conservative	62;	Mismatches 125; Indels 28; Gaps 4;
Qy	230	NARFWSLEDPYLYPL---KVELEKDEYTLDIGIRTISWDEKRLYLNGKPVFLKFGKHE	285	
Db	5	NPHLMQGEGLYELLYTAKSRTCEDIYPLRVGIRSAVAKGEQFLNHKPFYTFGFGHE	64	
Qy	286	EFPVLGQGTFFPLMIKDFNLKWINANSFRTSHYPPYSEEWLADRLGILVIDEAPHVGI	345	
Db	65	DADLRGKGFNDVLMVHDHALMDWIGANSYRTSHYPAEMLDWADEHGIVIDETAAGVF	124	
Qy	346	-----TRYHNPETOKIAEDNIRRMIDRHKNHPSVIMSVANEPESN	387	
Db	125	NLSLIGTFEAGNKKPKELYSEAVNGETQAHLQAIKELIARDKNHPSVVMWSIANEPDTR	184	
Qy	388	HPDAEGFFKALYETANEMDRTRPVVMVSM--DAPDETRDVALKYFDIVCVNRYGWI	445	
Db	185	PGGAREYFAPLAEATRKLDPTRPITCVNMFCDAHTDTISDL---FDVLCNRYGYWYV	240	
Qy	446	YQGRIEEGLQALEKDIEELYARHKRPIFVTEFGADAIAGIHYPDPQMFSEEQALVEKT	505	
Db	241	QSGDLTAEKVLLEKELLAWQEKLHQPIITTEYGVDTLAGLHSMYTDMMSEEYQCAWDMY	300	
Qy	506	IRLLLKDDYIIGTHVWAFADKTPQNVRRPILNHKGVFTDRQPKLVANVLRRLWS	561	
Db	301	HRVFDRAVSAVGEQVWNPADFATISQILRVGKNKGIFTRDRKPKSAAPLLQKRWT	356	

RESULT 9
D95842
probable beta-galactosidase (EC 3.2.1.23) [imported] - Sinorhizobium meliloti (strain 10
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C:Accession: D95842
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: D95842
A:Status: preliminary

A/Molecule type: DNA
A/Residues: 1-755 <KUR>
A/Cross-references: UNIPROT:Q92XF7; GB:AL591985; PIDN:CAC48404.1; PID:gl5139876; GSPDB:C
A/Experimental source: strain 1021, megaplasmid psymb
R;Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaule, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Weiss, D.H.; Wong, K.; Yeh, K.
A/Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A/Reference number: A96039; MUID:21368234; PMID:11474104
A/Contents: annotation
C/Genetics:
A/Gene: lacZ1; SMD21655
A/Genome: plasmid
C/Superfamily: beta-galactosidase
C/Keywords: glycosidase; hydrolase

Query Match	16.2%;	Score 487;	DB 2;	Length 755;
Best Local Similarity	25.7%;	Pred. No. 2.5e-25;		
Matches	155;	Conservative 90;	Mismatches 193;	Indels 164; Gaps 24;
QY	28	RPIAVGSMNE---QYQDLCEYEGDPYKTTFFVVPKLSQKHRLYFAAVNTDCEVFLN	83	
Db	30	QPISLPHNAVELPFNYFDERCYQRA-FTYQRLAWRPDPSGREVLVFDAMADAVVYLN	88	
QY	84	GEKVENHIEYLPFEVDVTGKVKSGENELRVVENRLKVGFPSPKVPDSTHTVGFSGSF	143	
Db	89	GEEIYAHRDGYTPFEARLLTDRLLEGDNLITVKID-----GSENPETLPFG-----	133	
QY	144	PPANPDFFPYGGIIRPVLIETDTHARLIDIVDTSE-----SEPE-----	183	
Db	134	--GRIDVLTGAGIYRVWLKVTDPVSIANIKIETRDVLSDTKAVSLRCDLSNPQGLSFSG	191	
QY	184	-----KGLGKVKKIEVSEAVGOEMTIKLGEEKKIRTSNRFVEGEFILLENARFWSLED	238	
Db	192	TISALLKNAAGEVLAEVAGETTGQSLAFEM-----DG---LKGLSLMDIDD	234	
QY	239	PVLYPLKVELE---KDEYTLDIGRTISWDEKRLYLNGKPVFLKGFKHEEFPVLGGGT	294	
Db	235	PVLVYIEVELRTGGSDCFPAHFGRTAFTTEGFRNLNGRPLKIRGLNRHQSFPVGYAM	294	
QY	295	FYPLMIKDFNLKWK-INANSFRTSHYPYSEEWLIDLRLGILVIDEAP--HVGITRVHY	350	
Db	295	GRTAQRDADIMKHLHCNLTVRTSHYPSQSKWFLDHCRIQLLVFEEIPGMQHIG-----	348	
QY	351	NPETOKIAEDNIRRMIDRHKNHPSVIMSV-ANEPESNHDPDAEGFEKALYETANEMDTR	409	
Db	349	GEEWKQEAIRNVRRMTERDWNHPSIYIWCVRINESQDSH---DFAETNRLARELDPTR	404	
QY	410	PVVWVSMMDAPDERTRDVALKFYDIVCVNRYYGWITYQRIEEGLQALEKDI---EELYA	466	
Db	405	-----QTGGVR-----YTIDSEFLIEDVYTMNDFILGNEELPG	436	
QY	467	RHRK-----PIFVTEFGADALAGIHYDPROMFSEEX-QAELVEKTIIRL	508	
Db	437	ANRPRTALRPQOECTGLPRKVPYLITEFG-----HMYPTKIYDQEQQAHEHVARHLE-	489	
QY	509	LKKDY---IIGTHWAFADFKTPQNVRRPILNHK-----GVFTRDRQPKLVA	553	
Db	490	VLNAAVGDPIGSAIGWCMFDYNT-----HKDFGSGDRICYHGVMDFRBPFAA	539	
QY	554	HV 555		
Db	540	YV 541		

RESULT 10
JU0275
beta-galactosidase (EC 3.2.1.23) - Thermoanaerobacterium thermosulfurigenes
C/Species: Thermoanaerobacterium thermosulfurigenes
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004

C/Accession: JU0275; JT0963
R;Burchardt, G.; Bahl, H.
Gene 106, 13-19, 1991
A/Title: Cloning and analysis of the beta-galactosidase-encoding gene from Clostridium t
A/Reference number: JU0275; MUID:92039055; PMID:1840542
A/Accession: JU0275
A/Molecule type: DNA
A/Residues: 1-716 <BUR>
A/Cross-references: UNIPROT:P26257; GB:M57579; NID:gl44838; PIDN:AAA23249.1; PID:gl44839
A/Experimental source: strain EM1
A/Accession: JT0963
A/Molecule type: protein
A/Residues: 1,'X',3-27 <BU2>
C/Genetics:
A/Gene: lacZ
C/Superfamily: beta-galactosidase
C/Keywords: glycosidase; homodimer; hydrolase
F;389,429,462/Active site: Glu, Tyr, Glu #status predicted

Query Match	16.1%;	Score 482;	DB 2;	Length 716;
Best Local Similarity	27.3%;	Pred. No. 5e-25;		
Matches	166;	Conservative 91;	Mismatches 201;	Indels 150; Gaps 27;
QY	30	IAVPGSMNEQYQDLCYEEG-----PFTY-----KTTFFVP	59	
Db	5	IPINNWW---YFKADYEEGEYKVDLRSFENVNLPHNTIELLYNYFDEKMYQIKSCYKYP	61	
QY	60	KKLSQKH---IRLYFAAVNTDCEVFLNGEKVGENHIEYLPFE--VDVTGKVKSGENELR	113	
Db	62	LHISEKYRDKVIYIHFEGVMAVQVYLNGLYIGEHKGGYTPDIRIDEVYDWKELNMLT	121	
QY	114	VVENRLKVGFPSPKVPDSTHTVGFSGSPRANPDFFPYGGIIRPVLIETDTHARLID	173	
Db	122	VYVDSTER-----SDIPPKG-----GQIDVLTGAGIYREVSLGIVDVFIKNI	164	
QY	174	WYDT-SESEPEKKL-----GKYVKIEVSEAVGOEMTIKLGEEKKIRTSN	219	
Db	165	KVETHGIYDNEKSLNLIYHLENLNHQSNGVKKYKINDKN-GKEVFYK-----EFN	214	
QY	220	RFVEG-----EFILEN--ARFWSLEDPVLYPLKVELE---KDEYTLDIGRTISWDEK	267	
Db	215	TYLDAVKDYVSFNIEENLKIDLMVDVDPNLYEIKVGMKINNPSDEYDNKFGFREAVFKPD	274	
QY	268	RLYLNGKPVFLKGFKHEEFPVLGGTFYPLMIKDFNLK-K-INANSFRTSHYPYSEEWL	326	
Db	275	GFTYLNKRLKLRGLNRHQSYPVGYAMPFRVQEKDAEILKNELHLNIVRTSHYPSQKHFL	334	
QY	327	DLADRLGILVIDEAPHVGITRVHYNPETOKIAEDNIRRMIDRHKNHPSVIMSV-ANEP	385	
Db	335	NKCEDELGLVFEELPG---WQYIGNSEWKVYAEQNLRREMITRDWNHPSIILWGVRI	391	
QY	386	SNHPDAEGFEKALYETANEMDTRPVVWVSMMDAPDERTRDVALKFYDIVCVNRYYGW	445	
Db	392	-----DDDAFYGMNKIAHEIDPTRQTGGVRYI-----TNSGFLE--DV-----YT	430	
QY	446	YQRIEGL-QALEKDIEELYARHKKPIFVTEFGADALAGIHYDPROMF-SEEQAEVLE	503	
Db	431	FNDFIHDGINFLRKQOEVLTGIEHNVPLYLTYEYNG-----HMYPTKRFDNERQMEHCL	484	
QY	504	KTIIRL---LKKDYIIGTHWAFADFKTPQNVRRPILNHK-----GVFTRDRQPK	549	
Db	485	RHLRIQNASYLDSDISGAIGWCAFDYNT-----HKDFGSGDRICYHGVMDFRBLP	534	
QY	550	KLVAAVLR 557		
Db	535	KFASVYVK 542		

RESULT 11
D86872
beta-galactosidase (EC 3.2.1.23) [imported] - Lactococcus lactis subsp. lactis (strain I
C/Species: Lactococcus lactis subsp. lactis
C/Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001

A;Reference number: A90981; MUID:84028567; PMID:6313347
A;Accession: A90981
A;Molecule type: DNA
A;Residues: 2-1024 <KAL>
A;Cross-references: UNIPROT:P00722; GB:V00296; NID:g41901; PIDN:CAA23573.1; PID:g1197203
A;Note: translation of initiator Met is not shown
R;Fowler, A.V.; Zabin, I.
J. Biol. Chem. 253, 5521-5525, 1978
A;Title: Amino acid sequence of beta-galactosidase. XI. Peptide ordering procedures and
A;Reference number: A92233; MUID:78218239; PMID:97298
A;Accession: A92233
A;Molecule type: protein
A;Residues: 2-1024 <FOW>
A;Note: this is the final paper in a series
R;Calos, M.P.; Miller, J.H.
Nature 285, 38-41, 1980
A;Title: Molecular consequences of deletion formation mediated by the transposon Tn9.
A;Reference number: A93224; MUID:80183189; PMID:6246435
A;Accession: A93224
A;Molecule type: DNA
A;Residues: 356-476 <CAL>
R;Rutshouser, E.C.; Richardson, J.P.
J. Mol. Biol. 208, 23-43, 1989
A;Title: Identification and characterization of transcription termination sites in the E
A;Reference number: S06878; MUID:89362462; PMID:2475637
A;Accession: S06878
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 1-147 <RUT>
A;Cross-references: EMBL:X16313; NID:g41903; PIDN:CAA34380.1; PID:g41904
R;Mikryukov, N.N.; Petrov, N.A.; Karginov, V.A.; Vassilenko, S.K.
Bioorg. Khim. 6, 1735-1736, 1980
A;Title: Nucleotide sequence of a lambda-plac 5-1 DNA region coding for a COOH-terminal
A;Reference number: I41218
A;Accession: I41218
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 967-971,'R',973-1022,'E',1024 <MIK>
A;Cross-references: GB:M38327; NID:g146061; PIDN:AAA23835.1; PID:g146062
R;Jacobson, R.H.; Zhang, X.; Dubose, R.F.; Matthews, B.W.
submitted to the Brookhaven Protein Data Bank, July 1994
A;Reference number: A65162; PDB:1BGL
A;Contents: annotation; X-ray crystallography, 2.50 angstroms, residues 4-1024
R;Jacobson, R.H.; Zhang, X.J.; Dubose, R.F.; Matthews, B.W.
Nature 369, 761-766, 1994
A;Title: Three-dimensional structure of beta-galactosidase from E. coli.
A;Reference number: A58594; MUID:94277211; PMID:8008071
A;Contents: annotation; X-ray crystallography, 2.50 angstroms
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: H64761
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1024 <BLAT>
A;Cross-references: GB:AE000141; GB:U00096; NID:g1786532; PIDN:AC73447.1; PID:g1786539;
A;Experimental source: strain K-12, substrain MG1655
R;Prenk, P.
Gene 122, 231-232, 1992
A;Title: Nucleotide sequence of the classical lacZ deletion delta M15.
A;Reference number: I40987; MUID:93083990; PMID:1339377
A;Accession: I40987
A;Molecule type: DNA
A;Residues: 1-11,43-50 <RES>
A;Cross-references: EMBL:X58252; NID:g40882; PIDN:CAA41206.1; PID:g40883
C;Genetics:
A;Gene: lacZ
A;Map position: 8 min
C;Complex: homotetramer
A;Description: catalyzes hydrolysis of lactose into galactose and glucose

C;Superfamily: beta-galactosidase
C;Keywords: glycosidase; homotetramer; hydrolase; magnesium
F;2-1024/Product: beta-galactosidase #status experimental <MAT>
F;2-50/Region: alpha complementation
F;51-218/Domain: 1, jelly-roll beta-barrel #status predicted <DM1>
F;219-334/Domain: 2, fibronectin type-III fold #status predicted <DM2>
F;335-627/Domain: 3, distorted TIM barrel #status predicted <DM3>
F;628-737/Domain: 4, fibronectin type-III fold #status predicted <DM4>
F;738-1024/Domain: 5, anti-parallel beta-sandwich #status predicted <DM5>
F;417,419,462/Binding site: magnesium (Glu, His, Glu) #status experimental
F;462,504,538/Active site: Glu, Tyr, Glu #status predicted
Query Match 13.8%; Score 414; DB 1; Length 1024;
Best Local Similarity 28.4%; Pred. No. 3.7e-20;
Matches 132; Conservative 59; Mismatches 172; Indels 102; Gaps 16;
QY 15 LNCVMNLEVTSSKDRPIAVPGSWNE-----QYQDLCYE--EG 48
DB 55 LNGEWRP--AMFPAPPAEAVPESWLECDLPEADTVVPPSNMQMHGYDAPITYTVTPITVNP 112
QY 49 PFT-----YKTFYVPKX-LSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEV 99
DB 113 PFVPTENPTGCTSLTFNVDESWLOEGQTRIFDGVNSAFHLMCNGRWGVGQDSRLPSEF 172
QY 100 DVTGKVKSGENELRVVVENRLKVGGFPSKVPDSTHTVGFSGSPPPANFDFPYGCIIRP 159
DB 173 DLSAFLRAGENRLAVMV-LRMSDGSY--LEDQ-----DWMRMSGIFRD 212
QY 160 VLIEFTDHARLIDIVWDTSESEPEKKLGKVKYKIEVSEBAVQ----EMTKLGEEBK 214
DB 213 VSLHKPTTQISDFHVATRENDPS----RAVLEAEVQMCBLRDYLRVTYSLWQGETQ 267
QY 215 IRTSNRFVEGEFI-----LENARFWSLEDPYLYPLKVELKEDEYTL---- 255
DB 268 VASGTAPFGEIIDERGGYADRVTLRLNVENPKLMSAEIPNLVRAVELHTADCTLIEAE 327
QY 256 --DIGRTISDEKRLYLNGKRVFLKFGKHEEFVPLGGCTFYPLMIKDFNLKWINANS 313
DB 328 ACDVGFEVRRIENGLLLNGKPLLRGVNRHEHPLHGQVNDQETMVQDILMKONNFNA 387
QY 314 FRTSHYPYSEEWLADRLGILVIDEA--PH--VGITRYHNPETOKIAEDNIRRMIDR 368
DB 388 VRCSHYPNHPLMYTLCDRYGLVYVDEANIEHGMVPMNRLTDDPRWLPAMSERVTRMVOR 447
QY 369 HKNHPSVIMWSVANBP--ESNHPDABEGFFKALYETANEMDRTRPV 411
DB 448 DRNHPSVILWSLGNESGHGANH-----DALYRWIKSVDSRPV 485
RESULT 14
E90678
beta-D-galactosidase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05099
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: E90678
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: E90678
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1024 <HAY>
A;Cross-references: UNIPROT:Q8X685; GB:BA000007; PIDN:BA833820.1; PID:g13359854; GSPDB:G
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: EC60397
C;Superfamily: beta-galactosidase
Query Match 13.5%; Score 405.5; DB 2; Length 1024;
Best Local Similarity 25.2%; Pred. No. 1.4e-19;
Matches 159; Conservative 88; Mismatches 216; Indels 169; Gaps 26;

```
QY 15 LINGVNLVTSKDRPIAVPGSWNE-----QYQDLCEY---EG 48
Db 55 LINGEW--QFVWFPAPAEAVPESWLECDLPDADTVVPSNMQMHGYDAPITYNTVPTVNP 112
QY 49 PFT-----YKTFYVPKX-LSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEEV 99
Db 113 PFVPTENPTGCSYLTFFNVDESWLQEGQTRIIIFDGVNSAFHLMCNGRWVGYQDSRLISEF 172
QY 100 DVTGKVKSGENELRVVVENRLKVGFPSPKVPDSTHTVGFSGFPANFDFFPYGCIIRP 159
Db 173 DLSAFLRAGENRLAVMV-LRWSDGSY---LEDQ-----DMWRMSGIFRD 212
QY 160 VLIEFTDHARILDIMWDTSESEPEKTL--GKVKVIEVSEAVQEMTIKLGSEKKIR 216
Db 213 VSLHKPTTQISDFHVATLFNDFSRVLEAEVQMYGELRDEL---RVTVSLMQGETQVA 269
QY 217 TSNRFVEGEFI-----LENARFWSLEDPLYPLKVELEKDEYTL----- 255
Db 270 SGTAPEGGEIIDERGGYADRVTTLGLNVENPKLMSAEIPNIYRAVVELHTADGTLIEAEAC 329
QY 256 DIGIRTIWDEKRLYLNGKPVFLKGFQKHEEPVYGQGTFFPLMKDFNLKWINANSFR 315
Db 330 DVGFREVRIENGLLLNGKPLLRGVNRHEHHPHGHQVMEQTMVQDILLMKQNNFNNAV 389
QY 316 TSHYPYSEEWLDLADRLGILVIDEA---PH--VGITRYHYNPETQKIAEDNIRRMIDRHK 370
Db 390 CSHYPNHPPLWYTLCDRYGLYVVDENANIEHGMVPMNRLLTDDPRWLPAMSERVTRMVQDR 449
QY 371 NHPSVIMMSVANEP--ESNHPDAEGFFKALYETANEMDRTRPV-----VMVS 415
Db 450 NHPSVIIWSLGNESGHGANH-----DALYRWIKSVDPSPVQYEGGADTSATDIICP 502
QY 416 NMDAPDERTRDVALKYFDIVCVNRYGWIYIOGRIEBGLQALEKDIEBLYARHHRKPIFVT 475
Db 503 MYARVDEDQPFPAVPKMSI-----KKWLSLPGEM-----RPLILC 537
QY 476 EFG---ADAIAGIHYDPQMFSEEYQA-----ELVEKTIRLLKKOYIIGTH 519
Db 538 EYAHAMGNSLGG-----FAKYWQAFRQYPRLOGGFVWDLVDQS---LIKTYD--NGN 584
QY 520 VMAF--ADFKTPQNVRRPILNHHKGVFTRDRQP 549
Db 585 PWSAYGDFGDTPNDRQFCMN--GLVFADRTF 614
```

RESULT 15

```
A85529
beta-D-galactosidase [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C/Accession: A85529
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: A85480; MUID:21074935; PMID:11206551
A/Accession: A85529
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1024 <STO>
A/Cross-references: UNIPROT:Q8X685; GB:AE005174; MID:G12513175; PIDN:AA654693.1; GSPDB:C
A/Experimental source: strain O157:H7, substrain EDL933
C/Genetics:
A/Gene: lacZ
C/Superfamily: beta-galactosidase
```

```
Query Match 13.5%; Score 405.5; DB 2; Length 1024;
Best Local Similarity 25.2%; Pred. No. 1.4e-19;
Matches 159; Conservative 88; Mismatches 216; Indels 169; Gaps 26;
QY 15 LINGVNLVTSKDRPIAVPGSWNE-----QYQDLCEY---EG 48
||| | : ||| | |
| : |
```

```
Db 55 LINGEW--QFVWFPAPAEAVPESWLECDLPDADTVVPSNMQMHGYDAPITYNTVPTVNP 112
QY 49 PFT-----YKTFYVPKX-LSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEEV 99
Db 113 PFVPTENPTGCSYLTFFNVDESWLQEGQTRIIIFDGVNSAFHLMCNGRWVGYQDSRLISEF 172
QY 100 DVTGKVKSGENELRVVVENRLKVGFPSPKVPDSTHTVGFSGFPANFDFFPYGCIIRP 159
Db 173 DLSAFLRAGENRLAVMV-LRWSDGSY---LEDQ-----DMWRMSGIFRD 212
QY 160 VLIEFTDHARILDIMWDTSESEPEKTL--GKVKVIEVSEAVQEMTIKLGSEKKIR 216
Db 213 VSLHKPTTQISDFHVATLFNDFSRVLEAEVQMYGELRDEL---RVTVSLMQGETQVA 269
QY 217 TSNRFVEGEFI-----LENARFWSLEDPLYPLKVELEKDEYTL----- 255
Db 270 SGTAPEGGEIIDERGGYADRVTTLGLNVENPKLMSAEIPNIYRAVVELHTADGTLIEAEAC 329
QY 256 DIGIRTIWDEKRLYLNGKPVFLKGFQKHEEPVYGQGTFFPLMKDFNLKWINANSFR 315
Db 330 DVGFREVRIENGLLLNGKPLLRGVNRHEHHPHGHQVMEQTMVQDILLMKQNNFNNAV 389
QY 316 TSHYPYSEEWLDLADRLGILVIDEA---PH--VGITRYHYNPETQKIAEDNIRRMIDRHK 370
Db 390 CSHYPNHPPLWYTLCDRYGLYVVDENANIEHGMVPMNRLLTDDPRWLPAMSERVTRMVQDR 449
QY 371 NHPSVIMMSVANEP--ESNHPDAEGFFKALYETANEMDRTRPV-----VMVS 415
Db 450 NHPSVIIWSLGNESGHGANH-----DALYRWIKSVDPSPVQYEGGADTSATDIICP 502
QY 416 NMDAPDERTRDVALKYFDIVCVNRYGWIYIOGRIEBGLQALEKDIEBLYARHHRKPIFVT 475
Db 503 MYARVDEDQPFPAVPKMSI-----KKWLSLPGEM-----RPLILC 537
QY 476 EFG---ADAIAGIHYDPQMFSEEYQA-----ELVEKTIRLLKKOYIIGTH 519
Db 538 EYAHAMGNSLGG-----FAKYWQAFRQYPRLOGGFVWDLVDQS---LIKTYD--NGN 584
QY 520 VMAF--ADFKTPQNVRRPILNHHKGVFTRDRQP 549
Db 585 PWSAYGDFGDTPNDRQFCMN--GLVFADRTF 614
```

Search completed: September 1, 2005, 19:33:26
Job time : 45 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2005, 19:14:24 ; Search time 177 Seconds
(without alignments)
1628.817 Million cell updates/sec

Title: US-09-936-759-6
Perfect score: 3001
Sequence: 1 MVRPQRNKKRFFILNGVMN.....TRDRQPKLVAVHVLRLMSEV 563

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2999	99.9	563	2 Q9X0F2	Q9X0F2 thermotoga
2	1011	33.7	570	2 Q97U11	Q97U11 sulfolobus
3	977.5	32.6	599	2 Q8XPI9	Q8XPI9 clostridium
4	972.5	32.4	599	2 Q8VNV4	Q8VNV4 clostridium
5	929	31.0	602	2 Q9AF42	Q9AF42 staphylococ
6	919.5	30.6	598	2 Q9AHJ8	Q9AHJ8 lactobacill
7	906	30.2	670	2 Q6NL66	Q6NL66 drosophila
8	905	30.2	656	2 Q9VBR0	Q9VBR0 drosophila
9	905	30.2	670	2 Q8MMB7	Q8MMB7 drosophila
10	902.5	30.1	648	1 BGLR_MOUSE	P12265 mus musculu
11	899.5	30.0	648	2 Q6IR10	Q6IR10 mus musculu
12	898	29.9	651	1 BGLR_FELCA	Q97524 felis silve
13	894.5	29.8	648	1 BGLR_RAT	P66760 ratu
14	888	29.6	651	1 BGLR_CANFA	Q18835 canis famil
15	882	29.4	603	2 Q93VT4	Q93VT4 arabidopsis
16	880	29.3	603	2 Q8FHA3	Q8FHA3 escherichia
17	879	29.3	603	1 BGLR_ECOLI	P05804 escherichia
18	879	29.3	603	1 BGLR_CEREA	Q9AHJ4 escherichia
19	872.5	29.1	648	1 BGLR_MOUSE	Q77695 cercopithec
20	865.5	28.8	651	1 BGLR_HUMAN	P08236 homo sapien
21	865	28.8	593	2 Q6A5C6	Q6A5C6 propionibac
22	862	28.7	628	2 Q95Q32	Q95Q32 caenorhabdi
23	861.5	28.7	603	2 Q6W7J7	Q6W7J7 ruminococcu
24	856.5	28.5	808	2 Q7TPJ3	Q7TPJ3 rattus norv
25	841.5	28.0	686	2 Q9V9T9	Q9V9T9 drosophila
26	826.5	27.5	599	2 Q8E6A6	Q8E6A6 streptococc
27	821.5	27.4	599	2 Q8E0N2	Q8E0N2 streptococc
28	811	27.0	567	2 Q7UCB6	Q7UCB6 shigella fl
29	809	27.0	567	2 Q83RC5	Q83RC5 shigella fl
30	728	24.3	660	2 Q7P2E2	Q7P2E2 anopheles g
31	691	23.0	368	2 Q8X671	Q8X671 escherichia

32	691	23.0	370	2 Q7ADL5	Q7ADL5 escherichia
33	531	17.7	459	2 Q8FMX0	Q8FMX0 corynebacte
34	487	16.2	755	2 Q92XF7	Q92XF7 rhizobium m
35	486.5	16.2	745	2 Q93IM0	Q93IM0 thermoanaer
36	482	16.1	716	1 BGAL_THETU	P26257 thermoanaer
37	481.5	16.0	998	1 BGAL_LACLA	Q48727 lactococcus
38	477.5	15.9	998	2 Q8VPJ3	Q8VPJ3 lactococcus
39	476.5	15.9	1085	2 Q85250	Q85250 thermotoga
40	472.5	15.7	996	2 Q87523	Q87523 lactococcus
41	460	15.3	743	1 BGAL_THET	P77989 thermoanaer
42	453	15.1	704	2 Q650P5	Q650P5 bacteroides
43	452.5	15.1	1084	1 BGAL_THEMA	Q56307 thermotoga
44	441.5	14.7	695	2 Q64ZE7	Q64ZE7 bacteroides
45	431.5	14.4	682	2 Q8A9Q8	Q8A9Q8 bacteroides

ALIGNMENTS

RESULT 1

ID	Q9X0F2	PRELIMINARY;	PRT;	563 AA.
AC	Q9X0F2;			
DT	01-NOV-1999 (TREMBlrel. 12, Created)			
DT	01-NOV-1999 (TREMBlrel. 12, Last sequence update)			
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)			
DE	Beta-glucuronidase.			
GN	OrderedLocusNames=TM1062;			
OS	Thermotoga maritima.			
OC	Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.			
OX	NCBI_TaxID=2336;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MSB8 / DSM 3109 / ATCC 43589;			
RX	MEDLINE=99287316; PubMed=10360571; DOI=10.1038/20601;			
RA	Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,			
RA	Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,			
RA	McDonald L.A., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,			
RA	Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.L.,			
RA	Heidelberg J.F., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,			
RA	Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;			
RT	"Evidence for lateral gene transfer between Archaea and Bacteria from			
RT	genome sequence of Thermotoga maritima.";			
RL	Nature 399:323-329 (1999).			
DR	EMBL; AE001766; AAD36143.1; -.			
DR	PIR; A72300; A72300.			
DR	HSSP; P08236; 1BHG.			
DR	TIGR; TM1062; -.			
DR	GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.			
DR	GO; GO:0005975; P:carbohydrate metabolism; IEA.			
DR	InterPro; IPR008979; Gal_bind_1ike.			
DR	InterPro; IPR006101; Glyco_hydro_2.			
DR	InterPro; IPR006102; Glyco_hydro_2ig.			
DR	InterPro; IPR006104; Glyco_hydro_2SB.			
DR	InterPro; IPR006103; Glyco_hydro_2TIM.			
DR	Pfam; PF00703; Glyco_hydro_2; 1.			
DR	Pfam; PF02836; Glyco_hydro_2_C; 1.			
DR	Pfam; PF02837; Glyco_hydro_2_N; 1.			
DR	PRINTS; PR00132; GLHYDRLASE2.			
DR	PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.			
KW	Complete proteome.			
SQ	SEQUENCE 563 AA; 65682 MM; 98C030B75D33B6C1 CRC64;			
Query Match				
Best Local Similarity 99.9%; Score 2999; DB 2; Length 563;				
Matches 562; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
QY 1 MVRPQRNKKRFFILNGVMNLVETSKDRPIAVPGSMNEQYQDLGYEGRPFTYKTFYVVK 60				
Db 1 MVRPQRNKKRFFILNGVMNLVETSKDRPIAVPGSMNEQYQDLGYEGRPFTYKTFYVVK 60				
QY 61 XLSQKIRILYFAAVVTDCEVFLNGEKGVENHIEYLPEFVDVTGKYKSGENELRVVVENRL 120				

Db 61 ELSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTGKVKSGENELRVVVENRL 120
QY 121 KVGFPSPKVPDSCGTHTVGFFGSPFPANPDFPPYGGIIRPVLIIEFTDHARILDIWDTSSES 180
Db 121 KVGFPSPKVPDSCGTHTVGFFGSPFPANPDFPPYGGIIRPVLIIEFTDHARILDIWDTSSES 180
QY 181 EPEKLGKVKVKEVSEEAAGQEMTIKLGEEBKIRTSNRFVEGEFIIENARFWSLEDPY 240
Db 181 EPEKLGKVKVKEVSEEAAGQEMTIKLGEEBKIRTSNRFVEGEFIIENARFWSLEDPY 240
QY 241 LYPLKVELEKDEYTLDIGIRTISWDEKRLYLNGKPVFLKGFGKHEFPVLGGGTYPPLMI 300
Db 241 LYPLKVELEKDEYTLDIGIRTISWDEKRLYLNGKPVFLKGFGKHEFPVLGGGTYPPLMI 300
QY 301 KDFNLLKWINANSFRTSHYPYSEEWLADRLGILVIDEAPHVGTIRYHNPETOKIAED 360
Db 301 KDFNLLKWINANSFRTSHYPYSEEWLADRLGILVIDEAPHVGTIRYHNPETOKIAED 360
QY 361 NIRMIDRHKNHPSVTIMSVANEPESNHPDAEGFFKALYETANEMDRTPVVMVSMMDAP 420
Db 361 NIRMIDRHKNHPSVTIMSVANEPESNHPDAEGFFKALYETANEMDRTPVVMVSMMDAP 420
QY 421 DERTRDVALKYFDIVCVNRYGWIYQGRIEEGLQALEKDIIEELYARHKRPIFVTEFGAD 480
Db 421 DERTRDVALKYFDIVCVNRYGWIYQGRIEEGLQALEKDIIEELYARHKRPIFVTEFGAD 480
QY 481 AIAGIHYDPQPMFSEEQALVEKTRILLKKDYIIGTHWAFADPKTPQNVRRPILNHK 540
Db 481 AIAGIHYDPQPMFSEEQALVEKTRILLKKDYIIGTHWAFADPKTPQNVRRPILNHK 540
QY 541 GVFTTRDQPKLVAVHVLRLMSEV 563
Db 541 GVFTTRDQPKLVAVHVLRLMSEV 563

RESULT 2

Q97U11 PRELIMINARY; PRT; 570 AA.
ID Q97U11
AC Q97U11
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Beta-glucuronidase (Gusb) (EC 3.2.1.31).
GN Name=gusb; OrderedlocusNames=SSO3036;
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726; DOI=10.1073/pnas.141222098;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Brauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006894; AAK43138.1; -.
DR PIR; C90485; C90485.
DR HSSP; P08236; 1BHG.
DR GO; GO:0004566, F:beta-glucuronidase activity; IEA.
DR GO; GO:0005975, P:carbohydrate metabolism; IEA.
DR InterPro; IPR008979; Gal bind like.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_21g.
DR InterPro; IPR006104; Glyco_hydro_25b.
DR InterPro; IPR006103; Glyco_hydro_21tm.
DR Pfam; PF00703; Glyco_hydro_2; 1.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.

DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLHYDLASE2.
KW Complete proteome; Glycosidase; Hydrolase.
SQ SEQUENCE 570 AA; 66795 MW; DEB2FEC8050AF189 CRC64;

Query Match 33.7%; Score 1011; DB 2; Length 570;
Best Local Similarity 39.2%; Pred. No. 5.2e-58;
Matches 230; Conservative 92; Mismatches 191; Indels 74; Gaps 14;

QY 15 LNCWNLAEVTSKDRP-----IAVGSWNEQYODLCYEEGPTTYKTTFFVYPK 60
Db 11 LGFWKFKIDMENTGEENGWYKGLSEDIYVPASWNEQNPXWDQFSGIAYQKDLFVSN 70
QY 61 XLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTGKVKSGENELRVVVENRL 120
Db 71 DNGNRKAMWVPEGAGYITKLTWNGEYGGTHGSGFTQKFPILKV---NEFNKIV--V 123
QY 121 KVGFPSPKVPDSCGTHTVGFFGSPFPAN-----FDFPPYGGIIRPVLIIEFTDHARILDIW 174
Db 124 KIDNTSPY-----NLPPARDLNNAAFDFFNYGIRPVYIEFVDECHVEDIT 171
QY 175 VDTSESEPEKLGKVKVKEVSEEAAGQEMTIKLGEEBKIRTSNRFVEGEFIIEN 230
Db 172 VYT-----KSYGHLKVEI-LSECNQRFSLRFLVDKEGRVILNESSNEVFEDK--VNN 222
QY 231 ARFWSLEDPYLYPLKVELE-----KDEYTLDIGIRTISWDEKRLYLNGKPVFLKGFGKHE 285
Db 223 VLPWSPDNPLYTLIVEMYVGNLKDVSVERIGFRDVEVKDGKTYLNGKPIFLKGFGRHE 282
QY 286 EFPVLGGGTFFPLMIKDFNLLKWINANSFRTSHYPYSEEWLADRLGILVIDEAP--HV 343
Db 283 DFPILGKFTYGAVLVDFYLMRKIGANSFRTSHYPYSENEHLADLADEMGFLVIEBPCLYS 342
QY 344 GITRYHYNPETQKI-----AEDNIRRMIDRHKNHPSVTIMSVANEPESNHPDAEGF 394
Db 343 NISRVMSQEEIAGKMGFDVKYFEKVRDTIKEMIRQHKRPSVTIMSVANEPSPDIREVAEF 402
QY 395 FKALYETANEMDRTRPVVMVSMMDAPDERTRDVALKYFDIVCVNRYGWIYQGRIEEGL 454
Db 403 IRREVELFKSLDSSRPVTFAS-----HRSVRDLALEYVDVISLNYHGWYTEMGIDSGV 457
QY 455 QALEKDIIEELYARH-RKPIFVTEFGADAIAGIHYDPQPMFSEEQALVEKTRILLKKD 513
Db 458 KVAIIELEEIHKKFPEKPIITTEFGADAIYGLHSDPQMWSEEQSEMIRKYLELREKD 517
QY 514 YIIGTHWAFADPKTPQNVRRPILNHKGVFTTRDQPKLVAVHVLRLM 560
Db 518 YIVGFHIWVAFDRTTPQNPSTRITLNRKGIFTTRDQPKLAAKAVBELF 564

RESULT 3

Q8XP19 PRELIMINARY; PRT; 599 AA.
ID Q8XP19
AC Q8XP19
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Beta-glucuronidase.
GN Name=bglr; OrderedlocusNames=CPE0147;
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13;
RX MEDLINE=21664373; PubMed=11792842; DOI=10.1073/pnas.022493799;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003185; BAB79853.1; -.

DR HSSP; P08236; 1BHG.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR Pfam; PF00703; Glyco_hydro_2; 1.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.
DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLHYDRLASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 599 AA; 68729 MW; 1AE4393869DAEF3B CRC64;

Query Match	32.6%;	Score 977.5;	DB 2;	Length 599;
Best Local Similarity	36.9%;	Pred. No. 8.7e-56;		
Matches 227;	Conservative 102;	Mismatches 210;	Indels 77;	Gaps 16;

Qy	1	MVRPQRNKKRFILLINGVWNLEVTISKDR-----PIAVPGSWE--QYODLC	44
Db	1	MLPYIITESRQLIDLSGIMKFWFKLNEGGLTEELSKAPLEDTIEMAVPSSYNDLVESQEV	60
Qy	45	YEEGPFTYKTFYVBPCKLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTGK	104
Db	61	DHVGVWVWYERNFTIPKTLNLERIVLRGSATHEAKVYLNCELLVEHKGFTPFEABINDL	120
Qy	105	VKSGENELRVVVENRLKVGFPSKVPDSCGHTVGFSGSP-----PANDFFPYGG	155
Db	121	LVSQDNRLTVAVNN-----IDETTLPGVLKVEYVDGKKVINKSVNFDFFNYAG	170
Qy	156	IIRPVLIETDARILDIWVDTSESEDEKLGKVKKIEVSEAVGQ-EMTIKLGEEKK	214
Db	171	IHRPVKIYTPKSYIEDITITVDFKENN---GYVNYEV---QAVGKCNIKVTIIDEENN	223
Qy	215	IRTSNRFVEGEFIBENARFWSLEDPYLPLKVELKEDEYTLD----IGIRTIWDEKRL	269
Db	224	IVAEGEGKEGKLTINNVLHMEPMNAVLYKLKVELLDDEIIDTYFEEFGVRTVEVKDGKF	283
Qy	270	YLNCKPVFLKFGFKHEEFVPLGGTFYPLMIKDFNLKMINANSFRTSHYPYSEEWLDA	329
Db	284	LINNKPFYFKGFGKHEDSYVNGRGINEALINIKDFNLMKMIGANSFRTSHYPYSEIIMRLA	343
Qy	330	DRGLGLVIDEAPHVGITRHYN-----PE-----TOKIAEDNIRMRIDRHKN	371
Db	344	DREGIVVIDETPAVGL---HLNFMATGFGGDAPKRDWTKEIGTKEAHERILRELVSBDKN	400
Qy	372	HPSVIMMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVMS-MMDAPDE-RTRDYA	428
Db	401	HPCVVMMSVANEPDSDSEGAKEYFEPLIKLTKELDPOKRPVTVTYLMSTPDRCKVGDIV	460
Qy	429	LKYFDIVCVNRYGWIYIQGRIEBCALKEKDIEELYARHRK-PIFVTEFGADAIAGIHY	487
Db	461	----DVLCLNRYGWIYVAGGDLEBAKMLEDELKGWEBCPKTPIMFTEYGADTVAGLHD	516
Qy	488	DPQOMFSEEQAEIVEXTIRLLKKDVIIGTHVWAPADFKTPQNVRRPILNHKGVFTDR	547
Db	517	TVPVMTFEEYQVEYYKANHEVMDCKNFVGEQVWNPADFATSGGIIRVOGNKKGIFTRE	576
Qy	548	QPKLVAVHLRRLWSEV	563
Db	577	KPKMIAHSLRERWTNI	592

RESULT 4		
Q8VNV4		
ID	Q8VNV4	PRELIMINARY; PRT; 599 AA.
AC	Q8VNV4;	
DT	01-MAR-2002 (TREMBLrel. 20, Created)	
DT	01-MAR-2002 (TREMBLrel. 20, last sequence update)	
DT	01-OCT-2003 (TREMBLrel. 25, last annotation update)	
DE	Beta-glucuronidase.	
OS	Clostridium perfringens.	
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;	
OC	Clostridium.	
OX	NCBI_TaxId=1502;	

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=13;
 RX MEDLINE=21945361; PubMed=11948145;
 RX DOI=10.1128/JB.184.9.2333-2343.2002;
 RA Briolat V., Reyssset G.;
 RT "Identification of Clostridium perfringens genes involved in the
 RT adaptive response to oxidative stress.";
 RL J. Bacteriol. 184:2333-2343(2002).
 DR EMBL; AJ420784; CAD12654.1; -.
 DR HSSP; P08236; 1BHG.
 DR GO; GO:004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:005975; P:carbohydrate metabolism; IEA.
 DR Pfam; PF00703; Glyco_hydro_2; 1.
 DR Pfam; PF02836; Glyco_hydro_2_C; 1.
 DR Pfam; PF02837; Glyco_hydro_2_N; 1.
 DR PRINTS; PR00132; GLHYDRIASE2.
 DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
 DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; UNKNOWN_1.
 SQ SEQUENCE 599 AA; 68715 MW; BB5386A4AAED48 CRC64;

Query Match	32.4%;	Score 972.5;	DB 2;	Length 599;
Best Local Similarity	36.7%;	Pred. No. 1.9e-55;		
Matches 226;	Conservative 102;	Mismatches 211;	Indels 77;	Gaps 16;

Qy	1	WVRPORNKKRFILLINGVNMLEVTSSKDR-----PIAVGGSWNE--QYODLC	44
Db	1	MLYPIITESROLIDLSGIWKFKLNEGNGLTEELSKAPLEDITIMAVPSSYNDLVESQEV	60
Qy	45	YEEGPFTYKTFYVVPKXLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGK	104
Db	61	DHVGWVWYERNFTIPKTLNRIYLRFGSATHEAKVYLLNGELLVEHKGFTPFEEAINDL	120
Qy	105	VKSGENELRVVVENRKLKVGFPKVPDSCGHTVGFSGFP-----PANDFPFPYGG	155
Db	121	LVSQDNRLTVAVN-----IIDETLPGVLKVEVVDGKVIKNSVNFDFFNAYAG	170
Qy	156	IIRPVLIETDHRILDIWVDTSESEPEKKLGKVKKIEVSEEAQG-EMTIKLGEEEK	214
Db	171	IHRPVKIYTTPKSYIEDITITVDFKERN--GYANVEV---QAVGCKNIKVTIIDEENN	223
Qy	215	IRTSNRFVEGEFIIENARFWSLEDPYLYPLKVELKDEYTLD-----IGIRTISWDEKRL	269
Db	224	IYAEGEKGEKGLTINNVLHMEPMNAYLLKLVELLDEEIIIDTYFEFEGVRTVEVKDGE	283
Qy	270	YLNKRPVFLKFGFKHEEFPVLGGQTFYPLMIKDFNLKWINANSFRTSHYPYSEEWLDA	329
Db	284	LINNPFYFKGFGKHEDSYVNGRGINEALINIKDFNLKMWIGANSFRTSHYPYSEEIWLA	343
Qy	330	DRLGILVIDEAPHVGITRYHYN-----PE-----TOKIAEDNIRRMIDRHKN	371
Db	344	DREGIVIDETPAVGL--HLNFMATGFGGDAPRKPDWKIEIGTKEAHERILRELVSBDKN	400
Qy	372	HPSYIMMSVANEPESNHPDAGCFKALYETANEMD-RTRPVVMS--WMDAPDE-RTRDVA	428
Db	401	HPCVMMSVANEPDSDSEGAKEYEPLIKLTKXELDPQKRPTVTVTYLMSTPDRCKVGDIV	460
Qy	429	LKXYDIVCVNRYYGWYIYQGRIEEGLQALEKDIIEELYARHRK-PIFTEFGADAIAGTHY	487
Db	461	----DVLCLNRYGWWYVAGDLEAKRMLDELEKGEERCPKTPIMEFTEYGADTVAGLHD	516
Qy	488	DEPQSESEYQAEIVEXTIRLLKKDYIIGTHVAFADFKTPQNVARRPILNHKGVPFTRDR	547
Db	517	TVPVMEFTEEYQVEYYKANHEVMDCKNFVGEQVWNFADFATSOGLIRVQGNKGIFTRER	576
Qy	548	QPKLVAHVLRRLMSEV	563
Db	577	KPKMIAHSLRERWTNI	592

RESULT 5
Q9AFA2
ID Q9AFA2 PRELIMINARY; PRT; 602 AA.

AC Q9AFA2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Beta-glucuronidase.
OS Staphylococcus sp. RLH1.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=156489;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RLH1;
RA Jefferson R.A., Kilian A., Keese P.K.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RLH1;
RA Kilian A.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF354044; AAK29422.1; -.
DR HSSP; P08236; 1BHG.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_2ig.
DR InterPro; IPR006104; Glyco_hydro_2SB.
DR InterPro; IPR006103; Glyco_hydro_2TIM.
DR Pfam; PF00703; Glyco_hydro_2_1. 1.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.
DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLYDRLASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
DR SEQUENCE 602 AA; 68701 MW; 312AFCD1634D577 CRC64;

Query Match	31.0%;	Score 929;	DB 2;	Length 602;
Best Local Similarity	35.6%;	Pred. No. 1.3e-52;		
Matches 221;	Conservative 93;	Mismatches 224;	Indels 82;	Gaps 15;

[illegible]

Db	459	VAEILDIVIALNRNGWYFDGGDLEAKVHLR--QEFHAMNKRCPGKPRIMITEYGADTVA	515
QY	484	GIHYDPQMFSEEQAELEVEKTRILLKDYIIGTHVMAFADFKTPQNVRRPILNHKGVF	543
Db	516	GFHDIDPVMFTEEQVEYYQANHVFDEFENVGEOAMNFADFATSGVMRVQGNKKGVF	575
QY	544	TRDRQPKLVAVLRLMSEV	563
Db	576	TRDRKPKLAHVFRERWTNI	595

RESULT 6

ID	Q9AHJ8	PRELIMINARY;	PRT;	598 AA.
AC	Q9AHJ8;			
DT	01-JUN-2001	(TREMBLrel. 17,		Created)
DT	01-JUN-2001	(TREMBLrel. 17,		Last sequence update)
DT	01-OCT-2003	(TREMBLrel. 25,		Last annotation update)
DE	Beta-glucuronidase.			
GN	Name=gusA;			
OS	Lactobacillus gasseri.			
OC	Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;			
OC	Lactobacillus.			
OX	NCBI_TaxId=1596;			

RP SEQUENCE FROM N.A.

RX MEDLINE=21141841; PubMed=11229918;
 RX DOI=10.1128/AEM.67.3.1253-1261.2001;
 RA Russell W.M.; Kleenhammer T.R.;
 RT "Identification and cloning of gusA, encoding a new beta-glucuronidase
 from *Lactobacillus gasseri* ADH.";
 RL Appl. Environ. Microbiol. 67:1253-1261 (2001).
 DR EMBL; AF305888; AAK07836.1; -.

DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_2ig.
DR InterPro; IPR006104; Glyco_hydro_2SB.
DR InterPro; IPR006103; Glyco_hydro_2TIM.
DR Pfam; PF00703; Glyco_hydro_2_1.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.
DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLHYDRLASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
SQ SEQUENCE 598 AA; 69762 MW; 5398F060B2DD887D CRC64;

Query Match	30.6%	Score 919.5;	DB 2;	Length 598;
Best Local Similarity	34.1%;	Pred. No. 5.6e-52;		
Matches 207;	Conservative 110;	Mismatches 227;	Indels 63;	Gaps 14;

```

QY      4 PORNKRREFILILNGVNLLEV-----SKDRP-----IAPGSMNE--QYODLCYEE 47
      7 PIONKRYRNTLNGTWQFETDPNSVGLDEGMNKELBDPEMPVPGTFAELTTKRDRKYTT 66
QY      48 GBFTYKTFYVPPKLLSQKHIRLYEAAVNTDCEVFLNGEKGVENHIEYLPFEVDVTGKVK 107
      67 GDFWYQKDFPISFLKKELYIRGCSVTHRAKVFINGHEVGQHEGGFLPFQVKISNYNY 126
QY      108 GE-NELRVVENRKLKVGFP---SKVPDSGTHTVGFGSFPANDFPEYGGIIRPVLI 163
      127 DQNRVTVLVNNELSEKAIPCGTEILDNQ-----KLAQPYDFEFNYSGIMRNWLL 179
QY      164 FTDHARILDIWDTSESEPEKKLGKVKKIEVSEEAVGQ-EMTIKLGEEKKIRTSNRFV 222
      180 ALPQSQITNFKLN-----YOLANNKATITYINIEANNNAEFKVTLPFNOKEVACATSKN 232
QY      223 EGEFILENAREFWSLEDPYLPKVELEK---DEYLLDIGIRTSIWDEKRLYLNGKPVF 277
      233 TSSLTIKNPHLWSPNDPYSYKIKIEMLEDGKTWDEYTDKIGIRTVKIVNDKILLNNHPIY 292
Db

```

QY	278	LKGFGEHFEFVLGGTFYPPLMIKDFMLLKMINANSFRTSHYPYSEEWLIDLADRLGILVI	337
		: : : : : : : : : :	
Db	293	LKGFGKHEDFNVLGKAVNESIIKRDYCMKMIGANCFRSSHYPYAEWYQYADKYGFLII	352
QY	338	DEAPHVIGITRYHYN-----PETOKIAEDNIRRMIDRHKNHPSVIMWS	379
		: : : : : : : : : :	
Db	353	DEVPAVGLNRSITNFLVNVTNSNQSHFPASKTVPPELKVHEQEIKEMIDRDQRHPSVIAMS	412
QY	380	VANEESNHPDAEGFFKALYETANEMD-RTRPVVMVSWMADPDERTRVALKYFDIVCVN	438
		: : : : : : : : : : : : :	
Db	413	LFNEBESTTOESYDYFKDIFAFARKLDPNRPYTGTGLVMGSPK--VDKLHPLCDFVCIN	470
QY	439	RYYGYMYIQG-RIEEOALEKDIEELY-ARHRKPIFYTEFGADALAGITHYDPOMFSSE	496
		: : : : : : : : : : :	
Db	471	RYYGYVAVAGPEIVNAKKMLEDELDCQNCLKNKPFVFTTEFGADTLSSSHRLPDMEWSOE	530
QY	497	YOAELEVETIRLLKKDYIIGHVMAFADFKTPONVARPILNHKGVFTRDRQPKLVAHYL	556
		: : : : : : : : : : :	
Db	531	YONEYYQMYPDIFKKYIPFIGELVMNPADFKTSEGIRVVGNDKGIFTRDREPKOIAFTL	590
QY	557	RRLMSEV 563	
		: : : :	
Db	591	KKRMOOL 597	

RESULT 7

Q6NL66		
ID	Q6NL66	PRELIMINARY; PRT; 670 AA.
AC	Q6NL66;	
DT	05-JUL-2004	(TREMBLrel. 27, Created)
DT*	05-JUL-2004	(TREMBLrel. 27, Last sequence update)
DT	05-JUL-2004	(TREMBLrel. 27, Last annotation update)
DE	RE15795p.	
GN	Name=CG15117;	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	

```

RC STRAIN=Berkeley;
RA Stapleton M., Carlson J., Chavez C., Frise E., George R., Pacleb J.,
RA Park S., Wan K., Yu C., Rubin G.M., Celniker S.;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BT012475; AAS93746.1; -.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR008979; Gal bind like.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_219.
DR InterPro; IPR006104; Glyco_hydro_25B.
DR InterPro; IPR006103; Glyco_hydro_2TIM.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF00703; Glyco_hydro_2; 1.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.
DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLHYDRLASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 670 AA; 77050 MW; 00AE0E67AE1D9E8C CRC64;

```

Query Match 30.2%; Score 906; DB 2; Length 670;
Best Local Similarity 34.7%; Pred. No. 5e-51;
Matches 212; Conservative 122; Mismatches 209; Indels 68; Gaps 17.

```

QY 1 MVRPQRNKRFFILINGVNL-----EVTSKDR---PIAVGSMNEQ 39
QY 2 LY:PRESE:REVRLSDGIWNFVRSDQANPTQGVDEWYAKELSKSRPTIPMPVBSYNDI 107
DB 48
QY 40 YQD-LCYEEGFTYKTFYVPKLSQ-KHIRLYFAVNTDCEVFLNGEKVGENHIEYLPF 97

```

Db	108	TTDNLRDHVGTVWYDRKFFVPRRWSKDQRIWLRFGSVHYEAYVWINGQKVKVHEMGLPF	167
QY	98	EVDVYTKVKSQ-ENELRVVVENRLKVGGF- ---SKVPDSGTHTVGGFSGFPANFDEFP	152
Db	168	EAEVTDLLSYGAENRITVMCDNALIQTVPRQGRITEVENDGMTI-----VQSYTFDFFN	222
QY	153	YGGIIRPVLIEFTDARILDIDWDTSESEPEKLGKVKVIEVSEEAVGQ-----EMTIK	207
Db	223	YAGIHRSVHLTYTPRTFIEEVEVTTNLSK-DATIGEYFYSVSVNGSANEADNVLQIQAN	281
QY	208	LGEEBKII---RTSNRFVEGEFIELENARFW-----SLEDPLYPLKVELEK-----DEX	253
Db	282	LYDXDGLVANATSDQKLGKLGQVNPVKPWWPYLMHSEPGYLQLEIKLATNDELDVY	341
QY	254	TLDIGIRTIWDEKRLYLNGKPVFLKGFGKHEEFPVLQGTFFPLMIKDFNLKWINANS	313
Db	342	RLTKGIRTLISWNSQFLINGKPVYFRGFGRHEDSDIRGGLDNALWVRDFNLKWIGANA	401
QY	314	FRTSHYPYSEEWLADRLGILVIDEAPHVGITRYHYNPETQKIAEDNIRRIDRHNHP	373
Db	402	YRISHYPYSEESMOPADEHGIMIDECPSVDTE--NFSQELLGKHKSLEQLIHRDRNHP	459
QY	374	SVIMSVANEPESNHPDAEGFFKALYETANEMDRTRPVVMSMDAPDERTRDVALKYFD	433
Db	460	SVVMSSTIANEPRTGSVSADSYFELVANFTRSLDKTRPITTAIAV---SNTQDKAGRSLD	515
QY	434	IVCVNRYYGWYIYQGRIEBGLQALEKDI EELYA---RHRKPIFVTEEGADAIAGIHYPD	490
Db	516	ISFNRYNAMYSNAGRLD---MITONVIDEAIAMNKRYNKPIIMSEYGADTLEGLHQPA	572
QY	491	QMESEEYQAEVLVEKTI RL--LLKKDYIIGTHVAFADFKTPQVNRPRPILNHKGVFTRDR	547
Db	573	YVWSEEFQTEVESRHFKA FDELKRKGWFIIGEFVWNPADFCTAQS YTRVGKNKGCVFTRAR	632
QY	548	QPKLVAVHLRR	558
Db	633	QPKAAHLLRK	643

RESULT 8

Q9V8R0		
ID	Q9V8R0	PRELIMINARY;
AC	Q9V8R0; Q8T0G7;	PRT; 656 AA.
DT	01-MAY-2000 (TREMBLrel. 13, Created)	
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)	
DT	25-OCT-2004 (TREMBLrel. 28, Last annotation update)	
DE	CG15117-PA (LD04718p) .	
GN	ORFNames=CG15117;	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	

[1] —
RN SEQUENCE FROM N.A.
RP MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazef R.G., Champe W., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

DT 01-OCT-2002 (TReMBLrel. 22, Created)

DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)

DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)

DE CG15117-PB.

GN ORFNames=CG15117;

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,

RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,

RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hosteln D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,

RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodagat, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,

RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=22426065; PubMed=12537568;

RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,

RA Patel S., Adams M., Champe M., Dugan S.P., Friese E., Hodgson A.,

RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,

RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,

RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,

RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;

RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila

RT melanogaster euchromatic genome sequence.";

RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=22426070; PubMed=12537573;

RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,

RA Patel S., Friese E., Wheeler D.A., Lewis S.E., Rubin G.M.,

RA Ashburner M., Celniker S.E.;

RT "The transposable elements of the Drosophila melanogaster euchromatin:

RT a genomics perspective.";

RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).

RN [4]

RP SEQUENCE FROM N.A.

RX MEDLINE=22426069; PubMed=12537572;

RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,

RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,

RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,

RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,

RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,

RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,

RA Lewis S.E.;

RT "Annotation of the Drosophila melanogaster euchromatic genome: a

RT systematic review.";

RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).

RN [5]

RP SEQUENCE FROM N.A.

RG FlyBase;

RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

RN [6]

RP SEQUENCE FROM N.A.

RG FlyBase;

RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE003797; AAM68434.1; -.

DR HSSP; P08236; 1BHG.

DR FlyBase; FBgn0034417; CG15117.

DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR InterPro; IPR008979; Gal_bind like.

DR InterPro; IPR006101; Glyco_hydro_2.

DR InterPro; IPR006102; Glyco_hydro_2ig.

DR InterPro; IPR006104; Glyco_hydro_2SB.

DR InterPro; IPR006103; Glyco_hydro_2TIM.

DR InterPro; IPR006025; Pept_M_zn_BS.

DR Pfam; PF00703; Glyco_hydro_2; 1.

DR Pfam; PF02836; Glyco_hydro_2_C; 1.

DR Pfam; PF02837; Glyco_hydro_2_N; 1.

DR PRINTS; PR00132; GLYDRLASE2.

DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.

DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.

DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.

SQ SEQUENCE 670 AA; 77036 MW; 7A9F23351E458C90 CRC64;

Query Match 30.2%; Score 905; DB 2; Length 670;

Best Local Similarity 34.7%; Pred. No. 5.8e-51;

Matches 212; Conservative 122; Mismatches 209; Indels 68; Gaps 17;

QY 1 MVRPQRNKKRFLILINGVWNL-----EVTSKDR---PIAVPGSNWQ 39

DB 48 MLYPRESETREVRSLDGIWNFVRSDQANPTQGVREWYAKELSKSRPTIPMPVPASYNDI 107

QY 40 YQD-LCYEEGPFTYKTFYVPKXLSQ-KHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPF 97

DB 108 TTDNLRDHVGTVYDRKFFVPRSNKSDQRIWLRFGSVHYEAYVWINGQKVVKHEMHLPF 167

QY 98 EVDVTGKVKSG-ENELRVVENRKLKVGFP----SKVPDSGTHYVFGSPFPANDFPP 152

DB 168 EAEVTDLLSYGAENRITVMCDNALIQTVPGGRITVEVPNDGCMTI-----VQSYTFDFFN 222

QY 153 YGGIIRPVLIIEFTDHARILDIWVDTSESEPEKLGKVKKIEVSEAVGQ-----EMTIK 207

DB 223 YAGIHRSVHLTYTPRTFIEVEVYTNLSK-DATVGEVYVSVSANAEADNVLOIQAN 281

QY 208 LGEEKKI---RTSNRFEVEGFLENARFW-----SLEDPLYLPKVELEK-----DEY 253

DB 282 LYDKDGLLVANATSDQKLGKQVNPVKPWPYLMHSEPGYLYQLEIKLATNDELIDVY 341

QY 254 TLDIGIRTISWDEKRLYLNGKPVFLKGFKGHEEPVLGGTFYPLMIKDFNLKWINANS 313

DB 342 RLKVGIRTLSSWNSQFLINGKPFYFRGFRHEDSDIRGKGLDNALMVRDFNLKIGANA 401

QY 314 FRSHYYPYSEEWLADRLGILVIDEAPHVGITRYHYNPETQKLAEDNIRMRIDRHKNHP 373

DB 402 YRTSHYYPYSESWFADHEGIMIDECPVDT--NFSQELLGKHKSSLEQLIHRDRNHP 459

QY 374 SVIMMSVANEPSNHPDAEGFFKALYETANEMDRTRPVVMVMSMDADPERTDYALKYFD 433

Db 460 SVVWMSIANEPRTGSVADSYSFELVANFTRSLDKTRPITAAIAV-----SNTQDKXGRSLD 515
Qy 434 IVCVNRYYGWIYQGRIEGLQALEKDIIEELEYA---RHKKPIFVTEFGADAIAGIHYDP 490
Db 516 IISFNRYNAWYSNAGRUD--MITQNVIDEAIAWNKRYNKPILIMSEYGADTLEGHMQPA 572
Qy 491 QMFSEEQVLAELVEKTRIRL--LLKKDYIIGTHVWAFADFKTPQNVRRPILNHKGVFTRDR 547
Db 573 YVWSEEFQTEVFSRHFKAPELRRKKGWIFGEFVWNFADFKAQSYTRVGNGKKGVFTRAR 632
Qy 548 QPKLVAHVLR 558
Db 633 QPKAAHLLRK 643
RESULT 10
BGLR_MOUSE
ID_BGLR_MOUSE STANDARD; PRT; 648 AA.
AC P12265; Q61601; Q64473; Q64474;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Beta-glucuronidase precursor (EC 3.2.1.31).
GN Name=Gusb; Synonyms=Gus, Gus-s;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88085188; PubMed=2891607;
RA Gallagher P.M., D'Amore M.A., Lund S.D., Elliott R.W., Pazik J.,
RA' Hohnan C., Korthagen T.R., Ganschow R.E.;
RT "DNA sequence variation within the beta-glucuronidase gene complex
among inbred strains of mice.";
RL Genomics 1:145-152(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88284700; PubMed=3397060;
RA Gallagher P.M., D'Amore M.A., Lund S.D., Ganschow R.E.;
RT "The complete nucleotide sequence of murine beta-glucuronidase mRNA
and its deduced polypeptide.";
RL Genomics 2:215-219(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89062453; PubMed=3196706;
RA D'Amore M.A., Gallagher P.M., Korthagen T.R., Ganschow R.E.;
RT "Complete sequence and organization of the murine beta-glucuronidase
gene.";
RL Biochemistry 27:7131-7140(1988).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/HeJ, and YBR; TISSUE=Sperm;
RX MEDLINE=89384641; PubMed=2779578;
RA Wawrzyniak C.J., Gallagher P.M., D'Amore M.A., Carter J.E., Lund S.D.,
RA Rinchik E.M., Ganschow R.E.;
RT "DNA determinants of structural and regulatory variation within the
murine beta-glucuronidase gene complex.";
RL Mol. Cell. Biol. 9:4074-4078(1989).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=88216590; PubMed=2835664;
RA Funkenstein B., Leary S.L., Stein J.C., Catterall J.F.;
RT "Genomic organization and sequence of the Gus-s alpha allele of the
murine beta-glucuronidase gene.";
RL Mol. Cell. Biol. 8:1160-1168(1988).
CC -1- FUNCTION: Plays an important role in the degradation of dermatan
and keratan sulfates.
CC -1- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = an alcohol +
D-glucuronate.
CC -1- SUBUNIT: Homotetramer.
CC -1- SUBCELLULAR LOCATION: Lysosomal.

CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 2 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03047; AAA37696.1; -.
DR EMBL; J02836; AAA98623.1; -.
DR EMBL; M63836; AAA63309.1; -.
DR EMBL; M28540; AAA63307.1; -.
DR EMBL; M28541; AAA63308.1; -.
DR EMBL; M19279; AAA37697.1; -.
DR PIR; A32576; A32576.
DR HSSP; P08236; 1BHG.
DR MGD; MGI:95872; Gus.
DR InterPro; IPR008979; Gal_bind_like.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_21g.
DR InterPro; IPR006104; Glyco_hydro_2SB.
DR InterPro; IPR006103; Glyco_hydro_2TIM.
DR Pfam; PF00703; Glyco_hydro_2; 1.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.
DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLHYDRLASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
KW Glycoprotein; Glycosidase; Hydrolase; Lysosome; Signal.
FT SIGNAL 1 22
FT CHAIN 23 648 Beta-glucuronidase.
FT ACT_SITE 447 447 Proton donor (By similarity).
FT CARBOHYD 172 172 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 416 416 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 591 591 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 627 627 N-linked (GlcNAc...) (Potential).
FT VARIANT 87 87 T -> I (in strain C3H/HeJ).
FT VARIANT 233 233 I -> T (in allele GUS-SA).
FT VARIANT 265 265 D -> G (in strain YBR and strain C3H/
HeJ).
FT VARIANT 320 320 V -> I (in strain YBR and strain C3H/
HeJ).
FT VARIANT 428 428 E -> K (in allele GUS-SA).
FT VARIANT 616 616 F -> L (in allele GUS-SA).
SQ SEQUENCE 648 AA; 74239 MW; 3D8C65ADB3B96D6 CRC64;
Query Match 30.1%; Score 902.5; DB 1; Length 648;
Best Local Similarity 35.7%; Pred. No. 8.1e-51;
Matches 223; Conservative 103; Mismatches 207; Indels 91; Gaps 21;
Qy 1 MVRPQRNKKRPILLNGVWNLGV-TSKDR-----PIAVGSWNEQ 39
Db 27 MLFPKSPSRRLKALDGLWHRADLSNNRLOGFEQWYRQPLRESGPVLDMPVPSSFNDI 86
Qy 40 YQDLCYEE--GPFTYKTFYVPKLSQ---KHIRLYFAAVNTDCEVFLNGEKVGENHIEY 94
Db 87 TQEAALRDFIGWVWYEREAIPRRWTQDTDMKRVLRINSADHYAAVWVNGIHVVEHGGH 146
Qy 95 LPFEVDVTGKYSK--ENELRVVVENRLKVGGFPSKVP-----DSGTHVGFSGSF 143
Db 147 LPFEADISKLVQSGPLTTCRTITAIINNTLT-----PHTLPQGTIVYKTDTSMPKGYF--V 200
Qy 144 PPANEDFPYGGIIRPVLIEFTDHARILDIWDTSESEPEKKLGKVK--VKIYSEEA VG 201
Db 201 QDTSFDFFNVAGLHRSVVLVTPPTTYIDITVITN---VEDQIGLVTYWISVGSEHF-- 255
Qy 202 QEMTIKLGESEKKIRTSNRFVEGEFILENARFW----SLSDP-YLYPLKV-----ELEKD 251
Db 256 -QLEVQLDDEGKVAHAGTQGOQLQVPSANLWMPYLMHHEHAYWYSLVAVKVTTSVTD 314
Qy 252 EYTLDIGRTISWDEKRLYLNGKPVFLKGFQGHIEFPVLGGGTFFYPLMIKDFNLKWINA 311


```
Db      315 YTTLPVGIRTVAVTKSKFLNGKPFYFGVKNKHEDSDIRGKGFDPWPLLVKDFNLRLWLG 374
QY      312 NSFRTSHYPYSEEWLADLRLGILVIDEAPHVGIT-----RYHNPETOKIAEDN 361
Db      375 NSFRTSHYPYSEEVQLCDRYGIVIDECEGVGIVLPQSGNESLRHHL-----EVMEE 429
QY      362 IRRMIDRHKHNPVIMSVANEPESNHPDAEGFFKALYETANEMDRTRPVVMVS--MMDA 419
Db      430 VR----RDKNHPAVVMWSVANEPSSALKPAAYYFKTLITHTKALDLTRPVTFSNAKYDA 485
QY      420 PDERTDVALKYFDIVCVNRYYGWYTYQGRIEBGLQALEKDIIBLYARHKRPIFVTEFGA 479
Db      486 -----DLGAPYVDVICVNSYFSWYHDYGHLEVIQPOLNSQFENWYKTHOKPIIQSEYGA 539
QY      480 DALAGIHYDPQPMFSEEQAEVLVEKTIKRL--LKDYIIGTHWAFADFKTPQNVRRPIL 537
Db      540 DAIPGHEDDPRMFSEEQAKAVLENYHSLVDQKRKEYVVGELIWNFADFMNTNQSPLRYIG 599
QY      538 NHKGVFTRDRQPKLVAVHLR-RLW 560
Db      600 NKKGIFTRQRPKTSAFILRERYW 623
```

RESULT 11

```
Q6IR10 ID Q6IR10 PRELIMINARY; PRT; 648 AA.
AC Q6IR10;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
D5 Glucuronidase, beta.
GN Name=Gusb;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NMRI; TISSUE=Mammary tumor;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heisler F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NMRI; TISSUE=Mammary tumor;
RA Director MGC Project;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC071226; AAH71226.1;
DR GO: GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro: IPR008979; Gal bind like.
DR InterPro: IPR006101; Glyco_hydro_2.
DR InterPro: IPR006102; Glyco_hydro_21g.
DR InterPro: IPR006104; Glyco_hydro_25b.
```

```
DR InterPro: IPR006103; Glyco_hydro_21m.
DR Pfam: PF00703; Glyco_hydro_2; 1.
DR Pfam: PF02836; Glyco_hydro_2_C; 1.
DR Pfam: PF02837; Glyco_hydro_2_N; 1.
DR PRINTS: PR00132; GLHYDRASE2.
DR PROSITE: PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE: PS00608; GLYCOSYL_HYDROL_F2_2; 1.
SQ SEQUENCE 648 AA; 74195 MW; CCB8F84C3CD6C498 CRC64;
```

Query Match 30.0%; Score 899.5; DB 2; Length 648;
Best Local Similarity 35.9%; Pred. No. 1.3e-50;
Matches 224; Conservative 101; Mismatches 208; Indels 91; Gaps 21;

```
QY      1 MVRPQNNKKRPFILINGVWNLEV-TSKDR-----PIAVPGSWNEQ 39
Db      27 MLFPKSPSRELKALDGLMHFRADLSNNRLQGFEQGWYRQPLRESGPVLDMPVPSSFNDI 86
QY      40 YQDLCYEE--GPFTYKTFYVPKLSQ--KHIRLYFAAVNTDCEVFLNGEKVGENHIEY 94
Db      87 TQEAALRDFIGWWMYEREAILPRRWTDPTDMRVILRINSAHYAAVAVWNGIHVEHEGGH 146
QY      95 LPFEVDVTGKVKSG---ENELRVVENRLKVGGFPSKVP-----DSGTHTVGFGSF 143
Db      147 LPFEADISKLVQSGPLTTCRITTAINTLT---PHTLPPTIYVKTDTSMYPKGYF--V 200
QY      144 PRANFDFFPYGGIIRPVLIETFDHARILDIWDTSESEPEKLGKVK--VKIEVSEAVG 201
Db      201 QDTSPDFENYAGLHRSVLYTPTTYIDITVITN---VEQDIGVTYWISVQSGEHF-- 255
QY      202 QEMTIKLGEEKKIRTSNRFEVEGFLEENARFW---SLEDP-YLYPLKV-----ELEKD 251
Db      256 -QLEVOQLDEGGKVAHGTGNQGLQVPSANLWPLYMHEHPAYMVSLEVKVTTTESVTD 314
QY      252 EYTLDIGIRTIWDEKRLYLNGKPYVLKFGKGHEEPVLGGQTFYPLMKDFNLKWINA 311
Db      315 YTTLPVGIRTVAVTKSKFLNGKPFYFGVKNKHEDSDIRGKGFDPWPLLVKDFNLRLWLG 374
QY      312 NSFRTSHYPYSEEWLADLRLGILVIDEAPHVGIT-----RYHNPETOKIAEDN 361
Db      375 NSFRTSHYPYSEEVQLCDRYGIVIDECEGVGIVLPQSGNESLRHHL-----EVMEE 429
QY      362 IRRMIDRHKHNPVIMSVANEPESNHPDAEGFFKALYETANEMDRTRPVVMVS--MMDA 419
Db      430 VR----RDKNHPAVVMWSVANEPSSALKPAAYYFKTLITHTKALDLTRPVTFSNAKYDA 485
QY      420 PDERTDVALKYFDIVCVNRYYGWYTYQGRIEBGLQALEKDIIBLYARHKRPIFVTEFGA 479
Db      486 -----DLGAPYVDVICVNSYFSWYHDYGHLEVIQPOLNSQFENWYKTHOKPIIQSEYGA 539
QY      480 DALAGIHYDPQPMFSEEQAEVLVEKTIKRL--LKDYIIGTHWAFADFKTPQNVRRPIL 537
Db      540 DAIPGHEDDPRMFSEEQAKAVLENYHSLVDQKRKEYVVGELIWNFADFMNTNQSPLRYIG 599
QY      538 NHKGVFTRDRQPKLVAVHLR-RLW 560
Db      600 NKKGIFTRQRPKTSAFILRERYW 623
```

RESULT 12
BGLR_FELCA
ID BGLR_FELCA STANDARD; PRT; 651 AA.
AC 097524;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Beta-glucuronidase precursor (EC 3.2.1.31).
GN Name=GUSB;
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT MPS VII LYS-351.

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M13962; AAA41228.1; -.
DR EMBL; Y00717; CA68705.1; -.
DR PIR; A25047; A25047.
DR HSSP; P08236; 1BHG.
DR RGD; 2772; Gueb.
DR InterPro; IPR008979; Gal bind like.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_2ig.
DR InterPro; IPR006104; Glyco_hydro_2SB.
DR InterPro; IPR006103; Glyco_hydro_2TIM.
DR Pfam; PF00703; Glyco_hydro_2; 1.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.
DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLHYDRLASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
KM Glycoprotein; Glycosidase; Hydrolase; Lysozyme; Signal.
FT SIGNAL 1 22
FT CHAIN 23 648
FT ACT_SITE 447 648
FT CARBOHYD 172 447
FT CARBOHYD 172 447
FT CARBOHYD 416 416
FT CARBOHYD 591 591
FT CARBOHYD 627 627
FT CONFLICT 14 14
FT CONFLICT 21 21
FT CONFLICT 487 487
SQ SEQUENCE 648 AA; 74793 MW; 5ADB8F5234F0907E CRC64;

Query Match 29.8%; Score 894.5; DB 1; Length 648;
Best Local Similarity 36.1%; Pred. No. 2.7e-50;
Matches 225; Conservative 98; Mismatches 210; Indels 91; Gaps 20;

QY 1 MVRPQRNKKRFLILNGVWN-----LEVTSKDRPIAVPGSWE- 38
Db 27 MLFPKETPSRELKVLGDLMSFRADYSNNRLQGFQKWYRQPLRESGPTLDMPPSSFNDI 86
QY 39 -QYODLCYEGRFTYKTFYVVPKXLSQKHIRLYFAVNTD---CEVFLNGEKVGENHIEY 94
Db 87 TQEAELRNFIQWVWYREAVLPQRWQDTDRRVVLRINSAHYAVVWVNGIHVEHGGH 146
QY 95 LPFEVDVTGKVKSG---ENELRVVENRLKVGFPKVP-----DSGHTVGFPGSF 143
Db 147 LPFEADITKLVQSGPLTFRVTIAINNTLT---PYTLRPGTIVYKTDPSMYPKGYF--V 200
QY 144 PRANDFRPYGGIIRPVLIETDHAFLIDIVDTSSEPEKLGKVKVKEVSEAVGQE 203
Db 201 QDISFDFFNYAGLHRSVLYTPTTYIDITVTT---DVDRDVGLVNVMISV-QGSDHFQ 256
QY 204 MTIKLGESEKKIRTSNRFEVEGEFILENARFW---SLEDP-YLYPLKV-----ELEKDEY 253
Db 257 LEVRLDDEDKIVARGTGNEGQLKVRALHLMWPLYLMEHPAYLYLSLEVTMTTPESVSDFY 316
QY 254 TLIDIGIRTISWDEKRLYLNGKRPVFLKGFQKHEEFVLGQGTFFPLMIKDFNLKMNANS 313
Db 317 TLPVGIRTVAVTKSKFLINGKRPFYFGVKNKHEDSDIRGRGFDWPLLIKDFNLRLMIGANS 376
QY 314 FRTSHYPYSEEWLDLADRLGILVIDEAPHVGIT-----RHYNPETOKIAEDNIR 363
Db 377 FRTSHYPYSEEVQLCDRXYIVIDECFVGIVLPSFGVSLRHL-----EVNDELVR 431
QY 364 RMIDRHKHPSVIMSVANEPESNHDAGEFFKALYETANEMDRTPRVVMVSMMDPDER 423
Db 432 ----RDKNHPAVVMVSVANEPVSSLKRPAGYFFKTLIAHTKALDPTRPVTVS-----N 480
QY 424 TR---DVALKFFDIYCVNRYGYWYIOGRIEEGLALEKDIIEELYARHKRPFVTEFGAD 480
Db 481 TRYDADMGAPYVDVICVNSYLSWYHDYGLLEVLITQLDLTSQFENMYKMYOKPIIOSEYCAD 540

QY 481 AIAGIHYPDPQMFSEERYQAEVVEKTRILL---KKDYIIGTHVMAFADFKTPQNVARPI 537
Db 541 AVSGLHEDPPRMFSEERYQTALLE-NYHLILDEKREYVIGELINRADFMTNOSPRLRVTG 599

QY 538 NHKGVFTRDRQPKLVAVHLR-RLW 560
Db 600 NKKGIFTRQNPKNMAFILRERYW 623

RESULT 14

RESULT 14
BGLR_CANFA ID BGLR_CANFA STANDARD; PRT; 651 AA.
AC 018835;
DT 15-JUL-1998 (Rel. 36, Last Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Beta-glucuronidase precursor (EC 3.2.1.31).
GN Name=GUSB;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_Taxid=9615;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT MPS VII HIS-166.
RX MEDLINE=98190525; Pubmed=9521879; DOI=10.1006/geno.1997.5189;
RA Ray J., Bouvet A., Desanto C., Pyfe J.C., Xu D., Wolfe J.H.,
RA Aguirre G.D., Patterson D.F., Haskins M.E., Henthorn P.S.;
RT "Cloning of the canine beta-glucuronidase cDNA, mutation
RT identification in canine MPS VII, and retroviral vector-mediated
RT correction of MPS VII cells.";
RL Genomics 48:248-253(1998).
CC -!- FUNCTION: Plays an important role in the degradation of dermatan
CC and keratan sulfates (By similarity).
CC -!- CATALYTIC ACTIVITY: A beta-D-glucuronoside + H(2)O = an alcohol +
CC D-glucuronate.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- DISEASE: Defects in GUSB are the cause of mucopolysaccharidosis
CC type VII (MPS VII), an inherited disease reported in humans, mice,
CC cats, and dogs.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 2 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF019759; AAC48809.1; -.
DR HSSP; P08236; 1BHG.
DR InterPro; IPR008979; Gal bind like.
DR InterPro; IPR006101; Glyco_hydro_2.
DR InterPro; IPR006102; Glyco_hydro_2ig.
DR InterPro; IPR006104; Glyco_hydro_2SB.
DR InterPro; IPR006103; Glyco_hydro_2TIM.
DR Pfam; PF00703; Glyco_hydro_2; 1.
DR Pfam; PF02836; Glyco_hydro_2_C; 1.
DR Pfam; PF02837; Glyco_hydro_2_N; 1.
DR PRINTS; PR00132; GLHYDRLASE2.
DR PROSITE; PS00719; GLYCOSYL_HYDROL_F2_1; 1.
DR PROSITE; PS00608; GLYCOSYL_HYDROL_F2_2; 1.
KM Disease mutation; Glycoprotein; Glycosidase; Hydrolase; Lysozyme;
KW Signal.
FT SIGNAL 1 22
FT CHAIN 23 651
FT ACT_SITE 450 450
FT CARBOHYD 172 450
FT CARBOHYD 419 419
FT CARBOHYD 630 630
FT VARIANT 166 166
By similarity.
Beta-glucuronidase.
Proton donor (By similarity).
N-linked (GlcNAc...) (Potential).
N-linked (GlcNAc...) (Potential).
N-linked (GlcNAc...) (Potential).
R -> H (in MPS VII; loss of activity).

Job time : 181 secs

•

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2005, 18:52:26 ; Search time 28 Seconds
(without alignments)
1500.979 Million cell updates/sec

Title: US-09-936-759-6
Perfect score: 3001
Sequence: 1 MVRPQRNKKRFFILLNGVMN.....TRDRQPKLVAAHVLRRLMSEV 563

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 segs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCtUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2999	99.9	563	US-09-270-957-6	Sequence 6, Appli
2	2999	99.9	563	US-09-270-957-21	Sequence 21, Appli
3	929	31.0	602	US-09-149-727-2	Sequence 2, Appli
4	929	31.0	602	US-09-270-957-2	Sequence 2, Appli
5	929	31.0	602	US-09-270-957-8	Sequence 8, Appli
6	929	31.0	602	US-09-270-957-15	Sequence 15, Appli
7	929	31.0	618	US-09-149-727-4	Sequence 4, Appli
8	925	30.8	607	US-09-149-727-8	Sequence 8, Appli
9	925	30.8	607	US-09-270-957-18	Sequence 18, Appli
10	925	30.8	615	US-09-270-957-28	Sequence 28, Appli
11	919.5	30.6	598	US-09-862-660-2	Sequence 2, Appli
12	899.5	30.0	648	US-09-715-858-4	Sequence 4, Appli
13	882	29.4	1010	US-09-118-276-12	Sequence 12, Appli
14	879	29.3	603	US-09-149-727-6	Sequence 6, Appli
15	879	29.3	603	US-09-270-957-17	Sequence 17, Appli
16	879	29.3	603	US-09-270-957-23	Sequence 23, Appli
17	877	29.2	832	US-08-630-820-7	Sequence 7, Appli
18	877	29.2	832	US-09-273-453-7	Sequence 7, Appli
19	875	29.2	604	US-09-893-525-37	Sequence 37, Appli
20	875	29.2	659	US-09-893-525-40	Sequence 40, Appli
21	875	29.2	850	US-09-893-525-42	Sequence 42, Appli
22	872.5	29.1	602	US-08-882-704A-5	Sequence 5, Appli
23	872.5	29.1	602	US-09-151-957-5	Sequence 5, Appli
24	872.5	29.1	602	US-09-151-957-5	Sequence 5, Appli
25	872.5	29.1	602	5432081-2	Patent No. 5432081
26	872.5	29.1	1242	US-09-488-270A-2	Sequence 2, Appli
27	865.5	28.8	613	US-09-149-727-5	Sequence 5, Appli

28	865.5	28.8	613	4	US-09-270-957-16	Sequence 16, Appli
29	865.5	28.8	651	4	US-09-715-858-2	Sequence 2, Appli
30	849.5	28.3	600	6	5268463-2	Patent No. 5268463
31	849.5	28.3	600	6	5268463-2	Patent No. 5268463
32	692.5	23.1	376	4	US-09-270-957-4	Sequence 4, Appli
33	692.5	23.1	376	4	US-09-270-957-19	Sequence 19, Appli
34	575	19.2	500	4	US-09-949-016-11697	Sequence 11697, A
35	574	19.1	372	4	US-09-270-957-3	Sequence 3, Appli
36	574	19.1	372	4	US-09-270-957-22	Sequence 22, Appli
37	496	16.5	540	4	US-09-270-957-5	Sequence 5, Appli
38	489.5	16.3	535	4	US-09-270-957-20	Sequence 20, Appli
39	416	13.9	1334	6	5476657-1	Patent No. 5476657
40	416	13.9	1334	6	5476657-1	Patent No. 5476657
41	414	13.8	1010	4	US-09-654-449-2	Sequence 2, Appli
42	414	13.8	1010	4	US-09-759-152A-2	Sequence 2, Appli
43	414	13.8	1121	1	US-07-789-915A-2	Sequence 2, Appli
44	414	13.8	1121	1	US-08-005-002C-2	Sequence 2, Appli
45	414	13.8	1121	1	US-08-487-203A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-270-957-6
; Sequence 6, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL-GLUCRONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Thermotoga maritima
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(563)
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-09-270-957-6

Query Match	99.9%	Score 2999;	DB 4;	Length 563;
Best local Similarity	100.0%;	Pred. No. 3.8e-256;		
Matches 563;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MVRPQRNKKRFFILLNGVMNLEVTSKDRPIAVPGSWNEQYODLCYBEGPPTYKTFYVPK	60	
DB	1	MVRPQRNKKRFFILLNGVMNLEVTSKDRPIAVPGSWNEQYODLCYBEGPPTYKTFYVPK	60	
QY	61	XLGQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTGKYKSGENELRVVVENRL	120	
DB	61	XLGQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTGKYKSGENELRVVVENRL	120	
QY	121	KVGGFSPSKVPDSCGTHTVGFFGSPFPANPDPFPYGGIIRPVLIEFTDHARILDIWDTSES	180	
DB	121	KVGGFSPSKVPDSCGTHTVGFFGSPFPANPDPFPYGGIIRPVLIEFTDHARILDIWDTSES	180	
QY	181	EPEKKLGKVKVKEVSEBAVGQEMTIKLGESEKIRTSNRFVEGGEFLENARFWSLBDPY	240	
DB	181	EPEKKLGKVKVKEVSEBAVGQEMTIKLGESEKIRTSNRFVEGGEFLENARFWSLBDPY	240	
QY	241	LYPLKVELEKDEYTDIGIRTSWDEKRLYLNGKRVFLKFGKHBEPVLGGGTFFPLMI	300	
DB	241	LYPLKVELEKDEYTDIGIRTSWDEKRLYLNGKRVFLKFGKHBEPVLGGGTFFPLMI	300	
QY	301	KDFNLKMINANSFRTSHYPYSEELDLADRLGILVIDEAPHVGTITRYHNPETOKIAED	360	
DB	301	KDFNLKMINANSFRTSHYPYSEELDLADRLGILVIDEAPHVGTITRYHNPETOKIAED	360	

Qy	361	NIRRMIDRHKNHPSVIMMSVANE	PESENHPDAEGFFKALYETANEMDRTRPVMVMSMDAP	420
Db	361	NIRRMIDRHKNHPSVIMMSVANE	PESENHPDAEGFFKALYETANEMDRTRPVMVMSMDAP	420
Qy	421	DERTRDVALKTFDIVCVNRRYYGWYIYQGRIBEGLOALEKDI	IBELYARHRKPIFVTEFGAD	480
Db	421	DERTRDVALKTFDIVCVNRRYYGWYIYQGRIBEGLOALEKDI	IBELYARHRKPIFVTEFGAD	480
Qy	481	AIAGIHYDPPQMFSEEQALVEKTRILLKKDYII	IGTHVWAFADFKTPQNVRRPILNHK	540
Db	481	AIAGIHYDPPQMFSEEQALVEKTRILLKKDYII	IGTHVWAFADFKTPQNVRRPILNHK	540
Qy	541	GVFTRDRQPKLVAVHLRLMSEV	563	
Db	541	GVFTRDRQPKLVAVHLRLMSEV	563	

RESULT 2

```

US-09-2/0957-21
; Sequence 21, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Thermotoga maritima
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(563)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-270-957-21

```

```
Query Match      99.9%; Score 2999; DB 4; Length 563;
Best Local Similarity 100.0%; Pred. No. 3.8e-256;
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	I	MVRPORNKKR	F	I	L	I	N	G	V	N	L	E	V	T	S	K	D	R	P	I	A	V	P	G	S	W	N	E	O	Y	O	D	L	C	Y	E	E	G	P	T	Y	K	T	T	F	Y	V	P	K		60	
Db	I	MVRPORNKKR	F	I	L	I	N	G	V	N	L	E	V	T	S	K	D	R	P	I	A	V	P	G	S	W	N	E	O	Y	O	D	L	C	Y	E	E	G	P	T	Y	K	T	T	F	Y	V	P	K		60	
QY	61	XLSQKHIRLY	F	A	A	V	N	T	D	C	E	V	F	L	N	G	E	K	V	G	E	N	H	I	E	Y	L	P	F	E	V	D	T	G	K	V	K	S	G	E	N	E	L	R	V	V	E	N	R	L		120
Db	61	XLSQKHIRLY	F	A	A	V	N	T	D	C	E	V	F	L	N	G	E	K	V	G	E	N	H	I	E	Y	L	P	F	E	V	D	T	G	K	V	K	S	G	E	N	E	L	R	V	V	E	N	R	L		120
QY	121	KVGGFSPSKVP	D	S	G	T	H	T	Y	G	F	F	G	S	F	P	P	A	N	D	F	F	P	Y	G	I	R	P	V	L	I	E	F	T	D	H	A	R	I	L	D	I	W	D	T	S	E	S		180		
Db	121	KVGGFSPSKVP	D	S	G	T	H	T	Y	G	F	F	G	S	F	P	P	A	N	D	F	F	P	Y	G	I	R	P	V	L	I	E	F	T	D	H	A	R	I	L	D	I	W	D	T	S	E	S		180		
QY	181	EPEKKLGKVKVK	I	E	V	S	E	E	A	V	G	O	E	M	T	I	K	L	G	E	E	E	K	I	R	T	S	N	R	F	V	E	G	E	F	I	L	E	N	A	R	F	W	S	L	E	D	P	Y		240	
Db	181	EPEKKLGKVKVK	I	E	V	S	E	E	A	V	G	O	E	M	T	I	K	L	G	E	E	E	K	I	R	T	S	N	R	F	V	E	G	E	F	I	L	E	N	A	R	F	W	S	L	E	D	P	Y		240	
QY	241	LYPLKVELEK	E	D	E	Y	T	L	D	I	G	I	R	T	I	S	W	D	E	K	R	L	Y	L	N	G	K	P	V	F	L	G	F	G	K	H	E	E	F	P	L	G	O	T	F	Y	P	L	M	I		300
Db	241	LYPLKVELEK	E	D	E	Y	T	L	D	I	G	I	R	T	I	S	W	D	E	K	R	L	Y	L	N	G	K	P	V	F	L	G	F	G	K	H	E	E	F	P	L	G	O	T	F	Y	P	L	M	I		300
QY	301	KDFNLLKMIN	A	N	S	F	R	T	S	H	Y	P	Y	S	E	W	L	D	L	A	D	R	L	G	I	L	V	I	D	E	A	P	H	V	G	I	T	R	Y	H	N	P	E	T	O	K	I	A	E	D		360
Db	301	KDFNLLKMIN	A	N	S	F	R	T	S	H	Y	P	Y	S	E	W	L	D	L	A	D	R	L	G	I	L	V	I	D	E	A	P	H	V	G	I	T	R	Y	H	N	P	E	T	O	K	I	A	E	D		360
QY	361	NIRRMIDRHKN	H	P	S	V	I	M	S	V	A	N	E	P	E	S	N	H	P	D	A	E	G	F	K	A	L	Y	E	T	A	N	E	M	D	R	T	R	P	V	M	S	M	D	A	P		420				
Db	361	NIRRMIDRHKN	H	P	S	V	I	M	S	V	A	N	E	P	E	S	N	H	P	D	A	E	G	F	K	A	L	Y	E	T	A	N	E	M	D	R	T	R	P	V	M	S	M	D	A	P		420				

[illegible]

RESULT 3

```

US-09-149-727-2
; Sequence 2, Application US/09149727
; Patent No. 6391547
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Kilian, Andrezej
; APPLICANT: Keese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 190106, 405
; CURRENT APPLICATION NUMBER: US/09/149, 727
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: US 60/058, 263
; EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-149-727-2

```

Query Match	31.0%;	Score 929;	DB 3;	Length 602;
Best Local Similarity	35.6%;	Pred. No. 2.5e-73;		
Matches 221; Conservative	93;	Mismatches 224;	Indels 82;	Gaps 15;

OY	I	MVRPQRNKKRFFILNGVMNLV-----TSKDRPIAVEGSWNE--QYODIC	44
Db	1	MLYPINTETRGVPDLNGVMNFKLIDYGKLGLEEKMYESKLTDTISMAVPSSSYNDIGVTKEIR	60
OY	45	YESCPFTYKTTPFYVPKXLSQKHRLYPAAVNTDCEVFLNGEKGVENHIEYLPEFVDVTGK	104
Db	61	NHGIVMYTEREFITVPAYLKDRIRIURFGSATHKAIYYVNGELVHEHKGGLPFPAEAINNS	120
OY	105	VKSGENELRVVVENRLKVGFPSKYVPDSGTHTVGFQS-----FPANEDFFPYG	154
Db	121	LROGMNRVTVAVDNIL-----DDSTLPVGLYSERHEEGLGKVI RNKNPNDFFNVA	170
OY	155	GIRPVLIEFTDHARILDIWDVTSESEPEKKLGVKVKIVESEA VGQEMTIKLG--EEE	212
Db	171	GIHRPVKIYTPTPTVEDISVATDENGR--TGTVTYVDFO--GRAETVKVSVDDEE	223
OY	213	KKIRTSNRFVEGEFILENARFWSLBEPYLYPLKVELEKDEYTLDI----GIRTISWDEK	267
Db	224	GKVVASTEGLSGNVEIPNVILMEPINTYLYQIKVELVNDGLTIDVYEBEPFGVRTVEVNDG	283
OY	268	RLYINGKPVFLKGGFKHEEPVLGGTFYPLMIKDNLWKIMANSFRTSHYPYSEEWLD	327
Db	284	KFLINNKPFYFKGGFKHEDTPINGRGFNEASNVMDEFNILKWI GANSFRTAHYPYSEELMR	343
OY	328	LADRLGILVIDEAPHVGITRYHN-----PETOKIA----EDNIRMIDR	368
Db	344	LADREGLVVIDETPAVGV--HLNMATTGLGEGSERVSTWEKIRTFEHQDVLRRELVSR	400
OY	369	HKNHPSVIMMSVANEPESNHDPAGEFPKALYETANEMD-RTRPVVMVSMMDAPDERTRDV	427
Db	401	DKNHPSVVMWSIANEAATEEGAGYEYFKPLVELTKELDPOKRPTVIVLFVMATPE--TDK	458
OY	428	ALKYFDIVCVNRRYGYIYOGRIBEGLOALEKDIIEELYARHR---KPIVFTEFGADAIA	483

```

Db      459 VAEILDIVIALNRNGWYFDGDLLEAAKVHLR---QEFHAMNKRCPGKPIMITTEYGADTVA 515
QY      484 GIHYDPQMFSEEQAEVLVEKTIIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHKGVF 543
Db      516 GFHDIDPVMFTEEQVEYYQANHVVDEFENFVGEQAMNFADFATSQGVMRVQGNKKGVF 575
QY      544 TRDRQPKLVAHVLRLMSEV 563
Db      576 TRDRKPKLAAHVFRERWTNI 595

```

```

RESULT 4
US-09-270-957-2
; Sequence 2, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-270-957-2

```

```

Query Match      31.0%; Score 929; DB 4; Length 602;
Best Local Similarity 35.6%; Pred. No. 2.5e-73;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

```

```

QY      1 MVRPQNRKKRFFILINGVWNLEV-----TSKDRPIAVPGSWNE--QYQDLC 44
Db      1 MLYPINTETRGVFDLNGVWNFKLDYKGLEEKWYESKLTDTISMAVPSSYNDIGVTKEIR 60
QY      45 YEEGPFTYKTFYVVPKLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPFEVDVTGK 104
Db      61 NHIGYVYEREFTPAYLKDQRIYLRFGSATHKAIYVNGELVVEHKGGLPFEAEINNS 120
QY      105 VKGENELRVVVENRLKVGFPSPKVPDSGTHTVGFFGS-----FPPANFDFPPYG 154
Db      121 LRDGMNRVTVAVDNIL-----DDSTLPVGLYSERHEEGLGKVIIRNKNPDEFENYA 170
QY      155 GIIRPVLIETDHAIRLDIWDTSSESEPEKLGKVKVKEIVSEAVQOEMTIKLG--EE 212
Db      171 GLHRPVKIYTPFTYVEDISVTDENGPR--TGIVTYTVDFQ---GKAETVKVSVVDEE 223
QY      213 KKIIRTSNRFEVEGEFIIENARFWSLEDPYLPKVELEKDEYTLDI-----GIRTSWDEK 267
Db      224 GKVAASTEGLSGNVEIPNVILMEPLNTLYLQIKVELVNDGLTIDVYEEPPGVRITVEVNDG 283
QY      268 RLYNGKPVFLKGFGKHEEPVLGOGTFYPLMIKDFNLKWINANSFRTSHYPYSEEWLD 327
Db      284 KFLINNKPFYFKGFGKHEDTPINGRGFNEASNVMDFNILKWIGANSFRTAHYPYSEELMR 343
QY      328 LADRLGILVIDEAPHVGITRYHYN-----PETOKIA-----EDNIRRMIDR 368
Db      344 LADREGLVVIDETPAVGV--HLNFMATTGLGEGSERVSTWEKIRTFEBHQDVLRELVS 400
QY      369 HKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVWVSMDAPDERTRDV 427
Db      401 DKNHPSVVMWSIANEAATEEBGAYEFKPLVELTKELDPQKRPTIIVLFVMATPE--TDK 458
QY      428 ALKYFDIVCVNRYGWIYIYQGRIEEGLQALEKDIIEELYARHR---KPIFVTEFGADAIA 483
Db      459 VAEILDIVIALNRNGWYFDGDLLEAAKVHLR---QEFHAMNKRCPGKPIMITTEYGADTVA 515
QY      484 GIHYDPQMFSEEQAEVLVEKTIIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHKGVF 543

```

```

Db      516 GFHDIDPVMFTEEQVEYYQANHVVDEFENFVGEQAMNFADFATSQGVMRVQGNKKGVF 575
QY      544 TRDRQPKLVAHVLRLMSEV 563
Db      576 TRDRKPKLAAHVFRERWTNI 595

```

```

RESULT 5
US-09-270-957-8
; Sequence 8, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-270-957-8

```

```

Query Match      31.0%; Score 929; DB 4; Length 602;
Best Local Similarity 35.6%; Pred. No. 2.5e-73;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

```

```

QY      1 MVRPQNRKKRFFILINGVWNLEV-----TSKDRPIAVPGSWNE--QYQDLC 44
Db      1 MLYPINTETRGVFDLNGVWNFKLDYKGLEEKWYESKLTDTISMAVPSSYNDIGVTKEIR 60
QY      45 YEEGPFTYKTFYVVPKLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPFEVDVTGK 104
Db      61 NHIGYVYEREFTPAYLKDQRIYLRFGSATHKAIYVNGELVVEHKGGLPFEAEINNS 120
QY      105 VKGENELRVVVENRLKVGFPSPKVPDSGTHTVGFFGS-----FPPANFDFPPYG 154
Db      121 LRDGMNRVTVAVDNIL-----DDSTLPVGLYSERHEEGLGKVIIRNKNPDEFENYA 170
QY      155 GIIRPVLIETDHAIRLDIWDTSSESEPEKLGKVKVKEIVSEAVQOEMTIKLG--EE 212
Db      171 GLHRPVKIYTPFTYVEDISVTDENGPR--TGIVTYTVDFQ---GKAETVKVSVVDEE 223
QY      213 KKIIRTSNRFEVEGEFIIENARFWSLEDPYLPKVELEKDEYTLDI-----GIRTSWDEK 267
Db      224 GKVAASTEGLSGNVEIPNVILMEPLNTLYLQIKVELVNDGLTIDVYEEPPGVRITVEVNDG 283
QY      268 RLYNGKPVFLKGFGKHEEPVLGOGTFYPLMIKDFNLKWINANSFRTSHYPYSEEWLD 327
Db      284 KFLINNKPFYFKGFGKHEDTPINGRGFNEASNVMDFNILKWIGANSFRTAHYPYSEELMR 343
QY      328 LADRLGILVIDEAPHVGITRYHYN-----PETOKIA-----EDNIRRMIDR 368
Db      344 LADREGLVVIDETPAVGV--HLNFMATTGLGEGSERVSTWEKIRTFEBHQDVLRELVS 400
QY      369 HKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVWVSMDAPDERTRDV 427
Db      401 DKNHPSVVMWSIANEAATEEBGAYEFKPLVELTKELDPQKRPTIIVLFVMATPE--TDK 458
QY      428 ALKYFDIVCVNRYGWIYIYQGRIEEGLQALEKDIIEELYARHR---KPIFVTEFGADAIA 483
Db      459 VAEILDIVIALNRNGWYFDGDLLEAAKVHLR---QEFHAMNKRCPGKPIMITTEYGADTVA 515
QY      484 GIHYDPQMFSEEQAEVLVEKTIIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHKGVF 543
Db      516 GFHDIDPVMFTEEQVEYYQANHVVDEFENFVGEQAMNFADFATSQGVMRVQGNKKGVF 575
QY      544 TRDRQPKLVAHVLRLMSEV 563
Db      576 TRDRKPKLAAHVFRERWTNI 595

```

```
RESULT 6
US-09-270-957-15
; Sequence 15, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-270-957-15
```

```
Query Match      31.0%; Score 929; DB 4; Length 602;
Best Local Similarity 35.6%; Pred. No. 2.5e-73;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY      1 MVRPQRNKKRFILILNGVWNLV-----TSKDRPIAVPGSWNE--QYQDLC 44
Db      1 MLVPINTETRGVFDLNGVWNPFLDYGKGLSEKWKYESKLTDTISMAVPSSYNDIGVTKEIR 60
QY      45 YEEGPFTYKTTFFVVPKLSQKHIRLYFAAVNTDCEVFLNGEKGXGHNHIEYLPFEVDVTGK 104
Db      61 NHIGVWYEREFTVPAYLKQRIVLRFGSATHKAIYVNGELVVEHKGGLPFEAEINNS 120
QY      105 VKSGENELRVVVENRLKVGSPSKVPDSCGTHVGFSG-----FPPANFDFFPYG 154
Db      121 LRDGMNRVTVAVDNLL-----DSTLPVGLYSERHEEGLGKVRNKNPFDFFNYA 170
QY      155 GIIRPVLIETDHAIRLDIWDVTSESEPEKLGKVKVKEIVSEEAVGQEMTIKLG--EE 212
Db      171 GLHRPVKIYTTPTFTYVEDISVTTDENGPR--TGTVTYTVDFQ---GKAETVKSVDDE 223
QY      213 KKIIRTSNRFVEGEFIELENARFWSLEDPYLYPLKVELEKDEYTLDI----GIRTIWDEK 267
Db      224 GKVVASTEGLSGNVEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYEEPPFGVRIVEVNDG 283
QY      268 RLYLNGKPVFLKFGKHEEPVLGQGTFFYPLMIKDFNLKWINANSFRTSHYPYSEEWLD 327
Db      284 KFLINNKPFYFKFGKHEDTPIINGRGFNEASNVMDFNILKWIGANSFRTAHYPSSELMR 343
QY      328 LADRLGILVIDEAPHVGITRYHYN-----PETOKIA-----EDNIRRMIDR 368
Db      344 LADREGLVVIDETPAVGV--HLNFMATTTGLGEGSERVSTWEKIRTFEHHQDVLRELVS 400
QY      369 HKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVMSMDAPDERTRDV 427
Db      401 DKNHPSVVMWSIANEAATEEGAYEYFKPLVELTKELDPQKRPTIVLFWMATPE--TDK 458
QY      428 ALKYFDIVCVNRYYGWYIYQGRIEEGLQALEKDIIEELYARHR---KPIFVTEFGADAIA 483
Db      459 VAEILIDVIALNRVNGWYFDGDDLAAKVHLR---QEFHAMNKRCPGKPIMITTEYGADTVA 515
QY      484 GIHYDPQPMFSESEYQAELEVEKTIIRLLKKDYIIGTHVWAFADFKTPQNVRRPILNHKGVF 543
Db      516 GFHDIDPVMFTEEYQVEYYQANHVVDFEFNFVGEQAWNFAFATSQGVMRVQGNKKGVF 575
QY      544 TRDRQPKLVAHVLRLMSEV 563
Db      576 TRDRKPKLAHVFRERWTNI 595
```

```
RESULT 7
US-09-149-727-4
; Sequence 4, Application US/09149727
```

```
; Patent No. 6391547
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Kilian, Andirzej
; APPLICANT: Reese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/09/149,727
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: US 60/058,263
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 618
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-149-727-4
```

```
Query Match      31.0%; Score 929; DB 3; Length 618;
Best Local Similarity 35.6%; Pred. No. 2.6e-73;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY      1 MVRPQRNKKRFILILNGVWNLV-----TSKDRPIAVPGSWNE--QYQDLC 44
Db      17 MLVPINTETRGVFDLNGVWNPFLDYGKGLSEKWKYESKLTDTISMAVPSSYNDIGVTKEIR 76
QY      45 YEEGPFTYKTTFFVVPKLSQKHIRLYFAAVNTDCEVFLNGEKGXGHNHIEYLPFEVDVTGK 104
Db      77 NHIGVWYEREFTVPAYLKQRIVLRFGSATHKAIYVNGELVVEHKGGLPFEAEINNS 136
QY      105 VKSGENELRVVVENRLKVGSPSKVPDSCGTHVGFSG-----FPPANFDFFPYG 154
Db      137 LRDGMNRVTVAVDNLL-----DSTLPVGLYSERHEEGLGKVRNKNPFDFFNYA 186
QY      155 GIIRPVLIETDHAIRLDIWDVTSESEPEKLGKVKVKEIVSEEAVGQEMTIKLG--EE 212
Db      187 GLHRPVKIYTTPTFTYVEDISVTTDENGPR--TGTVTYTVDFQ---GKAETVKSVDDE 239
QY      213 KKIIRTSNRFVEGEFIELENARFWSLEDPYLYPLKVELEKDEYTLDI----GIRTIWDEK 267
Db      240 GKVVASTEGLSGNVEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYEEPPFGVRIVEVNDG 299
QY      268 RLYLNGKPVFLKFGKHEEPVLGQGTFFYPLMIKDFNLKWINANSFRTSHYPYSEEWLD 327
Db      300 KFLINNKPFYFKFGKHEDTPIINGRGFNEASNVMDFNILKWIGANSFRTAHYPSSELMR 359
QY      328 LADRLGILVIDEAPHVGITRYHYN-----PETOKIA-----EDNIRRMIDR 368
Db      360 LADREGLVVIDETPAVGV--HLNFMATTTGLGEGSERVSTWEKIRTFEHHQDVLRELVS 416
QY      369 HKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVMSMDAPDERTRDV 427
Db      417 DKNHPSVVMWSIANEAATEEGAYEYFKPLVELTKELDPQKRPTIVLFWMATPE--TDK 474
QY      428 ALKYFDIVCVNRYYGWYIYQGRIEEGLQALEKDIIEELYARHR---KPIFVTEFGADAIA 483
Db      475 VAEILIDVIALNRVNGWYFDGDDLAAKVHLR---QEFHAMNKRCPGKPIMITTEYGADTVA 531
QY      484 GIHYDPQPMFSESEYQAELEVEKTIIRLLKKDYIIGTHVWAFADFKTPQNVRRPILNHKGVF 543
Db      532 GFHDIDPVMFTEEYQVEYYQANHVVDFEFNFVGEQAWNFAFATSQGVMRVQGNKKGVF 591
QY      544 TRDRQPKLVAHVLRLMSEV 563
Db      592 TRDRKPKLAHVFRERWTNI 611
```

```
RESULT 8
US-09-149-727-8
; Sequence 8, Application US/09149727
; Patent No. 6391547
```

```

; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Kilian, Andrzej
; APPLICANT: Keese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/09/149,727
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: US 60/058,263
; EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 8
; LENGTH: 607
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-149-727-8

```

```

Query Match      30.8%; Score 925; DB 3; Length 607;
Best Local Similarity 35.7%; Pred. No. 5.7e-73;
Matches 220; Conservative 92; Mismatches 223; Indels 82; Gaps 15;

```

```

QY 4 PQRNKKRFFILINGVWNLEV-----TSKDRPIAVPGSWE--QYODLCYEE 47
   | : | : | | | | | : : : : | : | : | : : :
Db 9 PINTETRGVFDLNGVWNFKLDYGKLEBKWYESKLTDTISMAVPSSTNDIGVTKEIRNHI 68

QY 48 GPFTYKTFYVPKLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEVDVTGKYS 107
   | : | : | | | : | : | : | : | : | : | : : :
Db 69 GYVWYEREFTVPAYLKQRIVLRFGSATHKAIYVNGELVVEHKGGLPFEAEINNSLRD 128

QY 108 GENELRVVENRLKVGSGPSPKVPDSGTHTVGFFGS-----PPRANFDFPYGGII 157
   | : | : | | : | : | : | : | : | : | : | :
Db 129 GMRNRTVAVDNIL-----DSTLPVGLYSERHEGLGKVI RNKPNDFFNAYAGLH 178

QY 158 RPLVLEFTDHARILDIVDTSESEPEKKLGKVKVIEVSEAVGQEMTIKLG--EEBKKI 215
   | | | | | : | : | : | : | : | : | : | :
Db 179 RPKIYTTPTTYVEDISVTDENG--TGTVTYVDFO---GKAETKVSVDDEGKV 231

QY 216 RTSNRFEVEGEFILENARFWSLEDPLYPLKVELKDEYTLDI----GIRTISWDEKRLY 270
   | : | : | : | : | : | : | : | : | : | :
Db 232 VASTEGLSGNVEIPNVILMEPLNTLYQIKVELVNDGLTIDVYEEPRGVRTVEVNDGKFL 291

QY 271 LNKGVFLKGFSGKHEEPVLGGTFYPLMIKDFNLKWINANSFRSHYYPYSEEWLAD 330
   | : | : | : | : | : | : | : | : | : | :
Db 292 INNKPFFYKGFSGKHEDTPINGRGFNEASNVDMFNILKWIIGANSFRTAHYYPYSEELMLAD 351

QY 331 RLGLVIDEAPHVGTIRYHYN-----PETOKIA-----EDNIRRMIDRHKN 371
   | : | : | | | | : | : | : | : | : | : | :
Db 352 REGLVVIDETPAVG--HLNFMAATGEGSERVSTWEKIRTFEHHQDVLRELVSBDKN 408

QY 372 HPSVIMSVANEPESNHPDAEGFFKALYETANEMD--RTRPVVMVSMMDAPDERTDVALK 430
   | | | : | : | : | : | : | : | : | : | :
Db 409 HPSVVMWSIANEAATEEGAYEYFKPLVELTKELDPQKRPTVILFVMATPE--TDKVAE 466

QY 431 YFDIVCVNRYYGWYIYQGRIEGLQALEKDIIEELYARHR---KPIFVTEFGADALAGIH 486
   | : | : | | | | : | : | : | : | : | :
Db 467 LIDVIALNRNGWYFDGDDLEAAKVHLR---QEFHAMNKRCPGKPMITEYGADTVAGFH 523

QY 487 YDPPQMFSEYQAEVLVEKTRILLKKDYIIGTHVMAFADFKTPQNVRRPILNHKGVTRD 546
   | | | : | : | | : : | : | : | : | : | :
Db 524 DIDPVMFTEEYQVEYQANHVVFDEFENFVGEQAMNADFATSGQVMRVQGNKKGVFTRD 583

QY 547 RQPKLVAHVLRRLWSEV 563
   | : | | | | | : | :
Db 584 RKPCLAAHVFRERWTNI 600

```

```

RESULT 9
US-09-270-957-18
; Sequence 18, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:

```

```

; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL -GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 607
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-270-957-18

```

```

Query Match      30.8%; Score 925; DB 4; Length 607;
Best Local Similarity 35.7%; Pred. No. 5.7e-73;
Matches 220; Conservative 92; Mismatches 223; Indels 82; Gaps 15;

```

```

QY 4 PQRNKKRFFILINGVWNLEV-----TSKDRPIAVPGSWE--QYODLCYEE 47
   | : | : | | | | | : : : | : | : | : : :
Db 9 PINTETRGVFDLNGVWNFKLDYGKLEBKWYESKLTDTISMAVPSSTNDIGVTKEIRNHI 68

QY 48 GPFTYKTFYVPKLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEVDVTGKYS 107
   | : | : | | | : | : | : | : | : | : | : : :
Db 69 GYVWYEREFTVPAYLKQRIVLRFGSATHKAIYVNGELVVEHKGGLPFEAEINNSLRD 128

QY 108 GENELRVVENRLKVGSGPSPKVPDSGTHTVGFFGS-----PPRANFDFPYGGII 157
   | : | : | | : | : | : | : | : | : | : | :
Db 129 GMRNRTVAVDNIL-----DSTLPVGLYSERHEGLGKVI RNKPNDFFNAYAGLH 178

QY 158 RPLVLEFTDHARILDIVDTSESEPEKKLGKVKVIEVSEAVGQEMTIKLG--EEBKKI 215
   | | | | | : | : | : | : | : | : | : | :
Db 179 RPKIYTTPTTYVEDISVTDENG--TGTVTYVDFO---GKAETKVSVDDEGKV 231

QY 216 RTSNRFEVEGEFILENARFWSLEDPLYPLKVELKDEYTLDI----GIRTISWDEKRLY 270
   | : | : | : | : | : | : | : | : | : | :
Db 232 VASTEGLSGNVEIPNVILMEPLNTLYQIKVELVNDGLTIDVYEEPRGVRTVEVNDGKFL 291

QY 271 LNKGVFLKGFSGKHEEPVLGGTFYPLMIKDFNLKWINANSFRSHYYPYSEEWLAD 330
   | : | : | : | : | : | : | : | : | : | :
Db 292 INNKPFFYKGFSGKHEDTPINGRGFNEASNVDMFNILKWIIGANSFRTAHYYPYSEELMLAD 351

QY 331 RLGLVIDEAPHVGTIRYHYN-----PETOKIA-----EDNIRRMIDRHKN 371
   | : | : | | | | : | : | : | : | : | : | :
Db 352 REGLVVIDETPAVG--HLNFMAATGEGSERVSTWEKIRTFEHHQDVLRELVSBDKN 408

QY 372 HPSVIMSVANEPESNHPDAEGFFKALYETANEMD--RTRPVVMVSMMDAPDERTDVALK 430
   | | | : | : | : | : | : | : | : | : | :
Db 409 HPSVVMWSIANEAATEEGAYEYFKPLVELTKELDPQKRPTVILFVMATPE--TDKVAE 466

QY 431 YFDIVCVNRYYGWYIYQGRIEGLQALEKDIIEELYARHR---KPIFVTEFGADALAGIH 486
   | : | : | | | | : | : | : | : | : | :
Db 467 LIDVIALNRNGWYFDGDDLEAAKVHLR---QEFHAMNKRCPGKPMITEYGADTVAGFH 523

QY 487 YDPPQMFSEYQAEVLVEKTRILLKKDYIIGTHVMAFADFKTPQNVRRPILNHKGVTRD 546
   | | | : | : | | : : | : | : | : | : | :
Db 524 DIDPVMFTEEYQVEYQANHVVFDEFENFVGEQAMNADFATSGQVMRVQGNKKGVFTRD 583

QY 547 RQPKLVAHVLRRLWSEV 563
   | : | | | | | : | :
Db 584 RKPCLAAHVFRERWTNI 600

```

```

RESULT 10
US-09-270-957-28
; Sequence 28, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL -GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957

```



```

; TYPE: PRT
; ORGANISM: Mus musculus
US-09-715-858-4

```

Query Match 30.0%; Score 899.5; DB 4; Length 648;
Best Local Similarity 35.9%; Pred. No. 1.1e-70;
Matches 224; Conservative 101; Mismatches 208; Indels 91; Gaps 21;

RESULT 13
US-09-118-276-12
Sequence 12, Application US/09118276
Patent No. 6693185
GENERAL INFORMATION:
APPLICANT: BABYCHUK, ELENA;
APPLICANT: KUSHNIR, SERGEI;
APPLICANT: DE BLOCK, MARC;
APPLICANT: INZE, DIRK
TITLE OF INVENTION: METHODS AND MEANS TO MODULATE PROGRAMMED
CELL DEATH IN EUKARYOTIC CELLS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: SIXBEY, FRIEDMAN, LEEDOM, & FERGUSON
STREET: 8180 GREENSBORO DRIVE, SUITE 800
CITY: MCLEAN,
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22102
COMPUTER READABLE FORM:

```

1 MEDIUM TYPE: 3-1/2" DISKETTE
2
3 COMPUTER: IBM-COMPATIBLE
4
5 OPERATING SYSTEM: MS-DOS
6
7 SOFTWARE: ASCII
8
9 CURRENT APPLICATION DATA:
10 APPLICATION NUMBER: US/09/118,276
11 FILING DATE: 17-JUL-1998
12
13 PRIOR APPLICATION DATA: NONE
14
15 ATTORNEY/AGENT INFORMATION:
16
17 NAME: SCHULMAN, ROBERT M.; SALEN, KENNETH H.
18
19 REGISTRATION NUMBER: 31,196; 43,077
20
21 REFERENCE/DOCKET NUMBER: 6201-0003
22
23 TELECOMMUNICATION INFORMATION:
24
25 TELEPHONE: (703) 790-9110
26
27 TELEFAX: (703) 883-0370
28
29 INFORMATION FOR SEQ ID NO: 12:
30
31 SEQUENCE CHARACTERISTICS:
32
33 LENGTH: 1010 RESIDUES
34
35 TYPE: AMINO ACID
36
37 STRANDEDNESS: SINGLE
38
39 TOPOLOGY: LINEAR
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
```

Query Match	29.4%;	Score 882;	DB 4;	Length 1010;
Best Local Similarity	34.0%;	Pred. No. 7.7e-69;		
Matches 206;	Conservative 109;	Mismatches 229;	Indels 62;	Gaps 11;

Qy	1	MVRPQRNKKRFLILINGWNLEV-----TSKDRPIAVPGSWMNQYODLCY	45
		: : : : : : : : : : : : :	
Db	408	MVRPVEPTPREIKLIDLGLMAFSLDRENGCIDQRWESALQESRAIAPGSGFNDQFADADI	467
Qy	46	EE--GPFTYKTFYVPVKXLSQKHRLYFAAVNTDCEVFLNGEKYGENHIEYLPFEVDVTG	103
		: : : : : : : : : : : : : : : : : : : : :	
Db	468	RNYAGNWMYQREVFIPKGMAGQRIVLRFDAVTHYGVWVNNQEWMEHQGGYTPFEADVTP	527
Qy	104	KVXSGEN-ELRVVENRLKVGSPSKV--PDSGTHTVGFFGSGPPANPDFPPYGGIIRP	159
		: : : : : : : : : : : : : : : : : : : : :	
Db	528	YIAGKSVRITVCVNNELNMQITPPGMVITDENGKKQSYF-----HDFPNYAGIHR	580
Qy	160	VLIEFTDHARLIDIVDWTSESEPEKKLGKVKVKEVSEAVGQEMTIKLGEEKKRITSN	219
		: : : : : : : : : : : : : : : : : : : : :	
Db	581	VMLYTPNTWDDITVTVTHVAQ---DCNHASVDMQV--VANGDVSVELRDADQOVATG	634
Qy	220	RFEVEGEFILENARFWSLEDPIYPL---KVELEKDEYTLDIGRTISWDEKRLYLNGKP	275
		: : : : : : : : : : : : : : : : : : : : :	
Db	635	QGTSGTLQVNVNPHLMQPGEGIYELCVTAKSQTECDIYPLRVGIRSAVAVKGEQFLINHP	694
Qy	276	VFLKGFGKHEEPFVLGGQTFYPLMIKDFNLKMTINANSFRTSHYPSSEWLDLADRIL	335
		: : : : : : : : : : : : : : : : : : : : :	
Db	695	FYFTGFGRHEDADLRGKGFNDLVMHDALMDMIGANSYRTSHIYAEEMLDWADENGIV	754
Qy	336	VIDEAPHVGI-----TRYHNPETOKIAEDNTRMIDRHKNHPSVIM	377
		: : : : : : : : : : : : : : : : :	
Db	755	VIDETAAVGFNLGLIGFEAGNKPKELYSEEAVENTEQAHLOAIKELIARDKNHPSVM	814
Qy	378	MSVANEPESNHPDAEGFFKALYETANEMDRTPVVMVSM--DAPDERTRDVALKYFDIV	435
		: : : : : : : : : : : : : : : : : : :	
Db	815	MSIANEPDTRPGARBYFAPLAETARKLDPTRPITCVNMFCDAHTDTISDL----FDVL	870
Qy	436	CVNRYGMYIYQGRIBEGLQALEKDI EELYARHRKPIFVTEFGADAIAGIHYPQMFSE	495
		: : : : : : : : : : : : : : : : : : :	
Db	871	CLNRYGMYVQSGDLETAKEVLEKELLAWQEKLHQPIITIEYGVDTLAGIHSMTVMSE	930
Qy	496	EYOAEIYVEKTIIRLLKCDYIIGTHWAFADPKTPQNVRRPILNHKGVFTRDRQPKLVAHV	555
		: : : : : : : : : : : : : : : : : : :	
Db	931	EYQCAWLDMYHVRVDRVSAVVGQVWNFADPATSQILRVGNNKGI FTRDRKPKSA AFL	990
Qy	556	LRRLMS 561	
		: :	
Db	991	LQKRW 996	

RESULT 14

```
US-09-149-727-6
; Sequence 6, Application US/09149727
; Patent No. 6391547
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Killion, Andrzej
; APPLICANT: Keese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/09/149,727
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: US 60/058,263
; EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-149-727-6
```

```
Query Match      29.3%; Score 879; DB 3; Length 603;
Best Local Similarity 33.8%; Pred. No. 6.5e-69;
Matches 205; Conservative 110; Mismatches 229; Indels 62; Gaps 11;

QY      1 MVRPQRNKKRFILINGVWNLEV-----TSKDRPIAVPGSWNEQYODLCY 45
      1 MLRPVETPTREIKKLDGLMAFSLDRENCGIDQRMWESALQESRAIAVPGSFNDQFADADI 60
Db
QY      46 EE--GPFTYKTFYVPKXLSQKHIRLYFAAVNTDCEVFLNGEKEVGENHIEYLPFEVDVTG 103
      61 RNVAAGNVWYQREVFIPKGMAGQRIYLRFDVATHYGKVVNNQEVMEHQGGYTPFEADVTP 120
Db
QY      104 KVKSGEN-ELRVVENRLKVGFPSPKV--PDSGTHTVGFFGSPFPANDFFPYGGIIRP 159
      121 YVIAGKSVRITVCVNNELNQITIPGMVITDENGKKQSYF-----HDFNYAGIHRS 173
Db
QY      160 VLIEFTDHARILDIWDTSESEPEKKLGKVKKIEVSEAVGQEMTIKLGEEKKIRTSN 219
      174 VMLYTPNTWDDITVTVTHVAQ---DCNHASVDWQV---VANGDVSVELRDADQOVVATG 227
Db
QY      220 RFVEGEFILENARFWSLEDPYLYPL---KVELEKDEYTLDIGIRTISWDEKRLYLNGKP 275
      228 QGTSGTLQVNPMLWQPEGGLYELCVTAKSQTECDIYPLRVGIRSAVKGEOFLINHKP 287
Db
QY      276 VFLKGFKGHEEPFVLGQGTFFPLMIKDFNLKMINANSFRSHYPYSEEWLIDLADRLGIL 335
      288 FYFTGFGRHEDADLRGKGFNDVLMVHDALMDWIGANSYRTSHYPYAEEMLDMADEHGIV 347
Db
QY      336 VIDEAPHVGI-----TRYHYNPETQKIAEDNIRRMIDRHKNHPSVIM 377
      348 VIDETAAVGFNLSLIGFEAGNKPKELYSEAVNGETQOAHLOAIKELIARDKNHPSVVM 407
Db
QY      378 WSVANEPESNHPDAEGFFKALYETANEMDRTRPVVWVSM--DAPDERTDVALKYFDIV 435
      408 WSIANEPDTRPQAGAREYFAPLAEATRKLDPTRPITCVNMVPCDAHTDTISDL----FDVL 463
Db
QY      436 CVNRYYGWYIYQGRIBEGLOALEKDIIEELYARHRKPIFVTEFGADAIAGIHYPDPQMFSE 495
      464 CLNRYYGWYVQSGDLETAEKVLEKELLAMQEKLHQPITITEYGVDTLAGLHSMYTDMMSE 523
Db
QY      496 EYQAEIYEKTIIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHKGVFTRDRQPKLVAAHV 555
      524 EYQCAWLDMYHRVFDVRSVAVVGEQVWNFADFATISQGITLRVGNNKKGIFTRDRKPKSA AFL 583
Db
QY      556 LRLRWS 561
      584 LQKRWT 589
Db
```

RESULT 15
US-09-270-957-17

```
; Sequence 17, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL -GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 17
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-270-957-17
```

```
Query Match      29.3%; Score 879; DB 4; Length 603;
Best Local Similarity 33.8%; Pred. No. 6.5e-69;
Matches 205; Conservative 110; Mismatches 229; Indels 62; Gaps 11;

QY      1 MVRPQRNKKRFILINGVWNLEV-----TSKDRPIAVPGSWNEQYODLCY 45
      1 MLRPVETPTREIKKLDGLMAFSLDRENCGIDQRMWESALQESRAIAVPGSFNDQFADADI 60
Db
QY      46 EE--GPFTYKTFYVPKXLSQKHIRLYFAAVNTDCEVFLNGEKEVGENHIEYLPFEVDVTG 103
      61 RNVAAGNVWYQREVFIPKGMAGQRIYLRFDVATHYGKVVNNQEVMEHQGGYTPFEADVTP 120
Db
QY      104 KVKSGEN-ELRVVENRLKVGFPSPKV--PDSGTHTVGFFGSPFPANDFFPYGGIIRP 159
      121 YVIAGKSVRITVCVNNELNQITIPGMVITDENGKKQSYF-----HDFNYAGIHRS 173
Db
QY      160 VLIEFTDHARILDIWDTSESEPEKKLGKVKKIEVSEAVGQEMTIKLGEEKKIRTSN 219
      174 VMLYTPNTWDDITVTVTHVAQ---DCNHASVDWQV---VANGDVSVELRDADQOVVATG 227
Db
QY      220 RFVEGEFILENARFWSLEDPYLYPL---KVELEKDEYTLDIGIRTISWDEKRLYLNGKP 275
      228 QGTSGTLQVNPMLWQPEGGLYELCVTAKSQTECDIYPLRVGIRSAVKGEOFLINHKP 287
Db
QY      276 VFLKGFKGHEEPFVLGQGTFFPLMIKDFNLKMINANSFRSHYPYSEEWLIDLADRLGIL 335
      288 FYFTGFGRHEDADLRGKGFNDVLMVHDALMDWIGANSYRTSHYPYAEEMLDMADEHGIV 347
Db
QY      336 VIDEAPHVGI-----TRYHYNPETQKIAEDNIRRMIDRHKNHPSVIM 377
      348 VIDETAAVGFNLSLIGFEAGNKPKELYSEAVNGETQOAHLOAIKELIARDKNHPSVVM 407
Db
QY      378 WSVANEPESNHPDAEGFFKALYETANEMDRTRPVVWVSM--DAPDERTDVALKYFDIV 435
      408 WSIANEPDTRPQAGAREYFAPLAEATRKLDPTRPITCVNMVPCDAHTDTISDL----FDVL 463
Db
QY      436 CVNRYYGWYIYQGRIBEGLOALEKDIIEELYARHRKPIFVTEFGADAIAGIHYPDPQMFSE 495
      464 CLNRYYGWYVQSGDLETAEKVLEKELLAMQEKLHQPITITEYGVDTLAGLHSMYTDMMSE 523
Db
QY      496 EYQAEIYEKTIIRLLKKDYIIGTHWAFADFKTPQNVRRPILNHKGVFTRDRQPKLVAAHV 555
      524 EYQCAWLDMYHRVFDVRSVAVVGEQVWNFADFATISQGITLRVGNNKKGIFTRDRKPKSA AFL 583
Db
QY      556 LRLRWS 561
      584 LQKRWT 589
Db
```

Search completed: September 1, 2005, 19:14:16
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 1, 2005, 19:04:53 ; Search time 167 Seconds
(without alignments)
1327.679 Million cell updates/sec

Title: US-09-936-759-6
Perfect score: 3001
Sequence: 1 MVRPQRNKKRFILINGVMN.....TRDRQPKLVAVHRLRMSEV 563

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1774312 seqs, 393823214 residues

Total number of hits satisfying chosen parameters: 1774312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep:*
- 18: /cgn2_6/ptodata/1/pubpaa/US10F_PUBCOMB.pep:*
- 19: /cgn2_6/ptodata/1/pubpaa/US10F_NEW_PUB.pep:*
- 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
- 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
- 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2999	99.9	563	15	US-10-364-649-6
2	2999	99.9	563	15	US-10-364-649-21
3	2999	99.9	563	15	US-10-369-493-2995
4	2999	99.9	563	18	US-10-757-093-20
5	1011	33.7	570	18	US-10-757-093-19
6	929	31.0	602	14	US-10-120-145-2
7	929	31.0	602	15	US-10-364-649-2
8	929	31.0	602	15	US-10-364-649-8
9	929	31.0	602	15	US-10-364-649-15
10	929	31.0	602	18	US-10-757-093-23
11	929	31.0	618	14	US-10-120-145-4

12	925	30.8	607	14	US-10-120-145-8	Sequence 8, Appli
13	925	30.8	607	15	US-10-364-649-18	Sequence 18, Appl
14	925	30.8	615	15	US-10-364-649-28	Sequence 28, Appl
15	920.5	30.7	641	18	US-10-757-093-2	Sequence 2, Appli
16	919.5	30.6	598	10	US-09-862-660-2	Sequence 2, Appli
17	919.5	30.6	598	15	US-10-673-935-2	Sequence 2, Appli
18	919.5	30.6	598	18	US-10-757-093-21	Sequence 21, Appl
19	905	30.2	670	18	US-10-757-093-12	Sequence 12, Appl
20	902.5	30.1	648	18	US-10-757-093-13	Sequence 13, Appl
21	899.5	30.0	648	15	US-10-421-175-4	Sequence 4, Appli
22	898	29.9	651	18	US-10-757-093-15	Sequence 15, Appl
23	894.5	29.8	648	18	US-10-757-093-14	Sequence 14, Appl
24	888	29.6	651	18	US-10-757-093-16	Sequence 16, Appl
25	883	29.4	634	18	US-10-757-093-4	Sequence 4, Appli
26	882	29.4	1010	9	US-09-118-276-12	Sequence 12, Appl
27	882	29.4	1010	16	US-10-705-197A-12	Sequence 12, Appl
28	879	29.3	603	14	US-10-161-403-106	Sequence 106, App
29	879	29.3	603	14	US-10-120-145-6	Sequence 6, Appli
30	879	29.3	603	15	US-10-364-649-17	Sequence 23, Appl
31	879	29.3	603	15	US-10-364-649-23	Sequence 842, App
32	879	29.3	603	15	US-10-369-493-842	Sequence 17, Appl
33	879	29.3	603	16	US-10-161-408-17	Sequence 106, App
34	879	29.3	603	18	US-10-757-093-22	Sequence 22, Appl
35	879	29.3	603	20	US-11-006-076-106	Sequence 28, Appl
36	877	29.2	618	14	US-10-356-088-28	Sequence 28, Appl
37	877	29.2	618	16	US-10-799-326-28	Sequence 7, Appli
38	877	29.2	832	17	US-10-632-815-7	Sequence 37, Appl
39	875	29.2	604	10	US-09-893-525-37	Sequence 37, Appl
40	875	29.2	604	16	US-10-763-380-37	Sequence 37, Appl
41	875	29.2	659	10	US-09-893-525-40	Sequence 40, Appl
42	875	29.2	659	16	US-10-763-380-40	Sequence 40, Appl
43	875	29.2	850	10	US-09-893-525-42	Sequence 42, Appl
44	875	29.2	850	16	US-10-763-380-42	Sequence 42, Appl
45	872.5	29.1	602	14	US-10-195-518-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-10-364-649-6
; Sequence 6, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL, B-GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Thermotoga maritima
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(563)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-364-649-6

Query Match 99.9%; Score 2999; DB 15; Length 563;
Best Local Similarity 100.0%; Pred. No. 1.9e-241;
Matches 563; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVRPQRNKKRFILINGVMNLEVTSTKDRPIAVPGSWNEQYQDLCYEAGPFTYKTFYVPK 60
Db 1 MVRPQRNKKRFILINGVMNLEVTSTKDRPIAVPGSWNEQYQDLCYEAGPFTYKTFYVPK 60

Qy	61	XL\$QKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKVKSGENELRVVENRL	120
Db	61	XL\$QKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPFEVDVTGKVKSGENELRVVENRL	120
Qy	121	KVGGFPSKVDPD\$GTHYVGF\$G\$FP\$P\$AN\$D\$FFPYGGIIRPVLI\$EFTD\$HARLID\$IW\$D\$T\$S\$E\$S	180
Db	121	KVGGFPSKVDPD\$GTHYVGF\$G\$FP\$P\$AN\$D\$FFPYGGIIRPVLI\$EFTD\$HARLID\$IW\$D\$T\$S\$E\$S	180
Qy	181	EPEKKLGKVKVKEI\$V\$E\$S\$E\$A\$V\$Q\$E\$M\$T\$IKL\$G\$E\$E\$K\$K\$IR\$T\$S\$N\$R\$F\$V\$E\$G\$E\$F\$IL\$E\$N\$A\$R\$F\$W\$S\$L\$E\$D\$P\$Y	240
Db	181	EPEKKLGKVKVKEI\$V\$E\$S\$E\$A\$V\$Q\$E\$M\$T\$IKL\$G\$E\$E\$K\$K\$IR\$T\$S\$N\$R\$F\$V\$E\$G\$E\$F\$IL\$E\$N\$A\$R\$F\$W\$S\$L\$E\$D\$P\$Y	240
Qy	241	LYPLKVELEKDEYTLDIGIRTISWDEKRLYLNGKRPVFLKGFGKHEEFVLCQGTFFPLMI	300
Db	241	LYPLKVELEKDEYTLDIGIRTISWDEKRLYLNGKRPVFLKGFGKHEEFVLCQGTFFPLMI	300
Qy	301	KDFNLLKWINANSFRTSHYPYSEEWLIDLADRLGILVIDEAPHVIGITRHYHNPETOKIAED	360
Db	301	KDFNLLKWINANSFRTSHYPYSEEWLIDLADRLGILVIDEAPHVIGITRHYHNPETOKIAED	360
Qy	361	NIRRMIDRHKNHPSVIM\$S\$V\$AN\$E\$P\$E\$S\$N\$H\$D\$A\$E\$G\$F\$K\$AL\$Y\$E\$T\$AN\$E\$M\$D\$R\$T\$R\$P\$V\$M\$S\$M\$D\$A\$P	420
Db	361	NIRRMIDRHKNHPSVIM\$S\$V\$AN\$E\$P\$E\$S\$N\$H\$D\$A\$E\$G\$F\$K\$AL\$Y\$E\$T\$AN\$E\$M\$D\$R\$T\$R\$P\$V\$M\$S\$M\$D\$A\$P	420
Qy	421	DERTRDVALKYFDIVCNRRYGYIYQGRIEEGLQALEKDI\$E\$E\$L\$Y\$A\$R\$H\$R\$K\$P\$I\$F\$V\$T\$E\$F\$G\$A\$D	480
Db	421	DERTRDVALKYFDIVCNRRYGYIYQGRIEEGLQALEKDI\$E\$E\$L\$Y\$A\$R\$H\$R\$K\$P\$I\$F\$V\$T\$E\$F\$G\$A\$D	480
Qy	481	AIAGIHYDPQOMFSE\$E\$Y\$Q\$A\$E\$L\$V\$E\$K\$T\$IR\$LL\$K\$D\$Y\$II\$G\$H\$W\$A\$F\$A\$D\$F\$K\$T\$P\$Q\$N\$V\$R\$P\$IL\$N\$H\$K	540
Db	481	AIAGIHYDPQOMFSE\$E\$Y\$Q\$A\$E\$L\$V\$E\$K\$T\$IR\$LL\$K\$D\$Y\$II\$G\$H\$W\$A\$F\$A\$D\$F\$K\$T\$P\$Q\$N\$V\$R\$P\$IL\$N\$H\$K	540
Qy	541	GVFTDRQPKLV\$A\$H\$V\$L\$R\$L\$M\$S\$E\$V\$563	
Db	541	GVFTDRQPKLV\$A\$H\$V\$L\$R\$L\$M\$S\$E\$V\$563	

```

RESULT 2
US-10-364-649-21
; Sequence 21, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Thermotoga maritima
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(563)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-364-649-21

```

Query Match	99.9%;	Score 2999;	DB 15;	Length 563;
Best Local Similarity	100.0%;	Pred. No. 1.9e-241;		
Matches 563;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1 MVRPQRNKKRFFILLINGVMNL EYTSKDRPIAVGSGWNEQYODLCYEEGPFTYKTFYVVPK 60
Db	1 MVRPQRNKKRFFILLINGVMNL EYTSKDRPIAVGSGWNEQYODLCYEEGPFTYKTFYVVPK 60

QY	61	XLSOKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLRPEVDVTGKVKSGENELRVVVENRL	120
Db	61	XLSOKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLRPEVDVTGKVKSGENELRVVVENRL	120
QY	121	KVGGFPSKVPDSCGTHTVGFFGSGFPANFDFPPYGGIIRPVLIIEFTDHARILDIWVDTSES	180
Db	121	KVGGFPSKVPDSCGTHTVGFFGSGFPANFDFPPYGGIIRPVLIIEFTDHARILDIWVDTSES	180
QY	181	EPEKKLGKVKVKIIEVSEEAVGQEMTIKLGEEEKKIRTSNRFVEGEFTLENARFWSLEDPY	240
Db	181	EPEKKLGKVKVKIIEVSEEAVGQEMTIKLGEEEKKIRTSNRFVEGEFTLENARFWSLEDPY	240
QY	241	LYPLKVELEKDEYTLDIGIRTISWDEKRLYLNGKRPVFLKGFGKHSEFPVLGGTFYPLMI	300
Db	241	LYPLKVELEKDEYTLDIGIRTISWDEKRLYLNGKRPVFLKGFGKHSEFPVLGGTFYPLMI	300
QY	301	KDFNLLKMINANSFRTSHYPYSEEWDLADRLGILYIDEAPHVGI TRHYNPETOKIAED	360
Db	301	KDFNLLKMINANSFRTSHYPYSEEWDLADRLGILYIDEAPHVGI TRHYNPETOKIAED	360
QY	361	NIRRMIDRHKNHPSVIMMSVANEPESNHPDAEGFFKALYETANEMDRTRPVMMVSMMDAP	420
Db	361	NIRRMIDRHKNHPSVIMMSVANEPESNHPDAEGFFKALYETANEMDRTRPVMMVSMMDAP	420
QY	421	DETRTRDVALKYFDIVCVNRRYYGWYIYQGRIEEGLQALEKDI EELYABHRKPI FVTEFGAD	480
Db	421	DETRTRDVALKYFDIVCVNRRYYGWYIYQGRIEEGLQALEKDI EELYABHRKPI FVTEFGAD	480
QY	481	AIAGIHYDPROMFSEEEYOAELEVETIRLLKKDYIIGTHVMAFADFPTQPONRRPILNHK	540
Db	481	AIAGIHYDPROMFSEEEYOAELEVETIRLLKKDYIIGTHVMAFADFPTQPONRRPILNHK	540
QY	541	GVFTRDRQPKLVAVHLRLMSEV	563
Db	541	GVFTRDRQPKLVAVHLRLMSEV	563

```

RESULT 3
US-10-369-493-2995
; Sequence 2995, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052) B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2995
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Thermotoga maritima
US-10-369-493-2995

```

Query Match	99.9%	Score 2999;	DB 15;	Length 563;
Best Local Similarity	99.8%;	Pred. No. 1.9e-241;		
Matches 562; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

Ox
1 MVRPQRNKKRIFILINGVNMIEVTSKDRPIAVPGSMNEQYQDLCEEGPEFTYKTFYVPK 60

Db
1 MVRPQRNKKRIFILINGVNMIEVTSKDRPIAVPGSMNEQYQDLCEEGPEFTYKTFYVPK 60

Qy	61	XL	S	K	H	I	R	L	Y	F	A	A	V	N	T	D	C	E	V	F	L	N	G	E	K	V	G	E	N	H	I	E	Y	L	P	E	V	D	T	G	K	V	K	S	G	E	N	E	L	R	V	V	E	N	R	L	120	
Db	61	EL	S	K	H	I	R	L	Y	F	A	A	V	N	T	D	C	E	V	F	L	N	G	E	K	V	G	E	N	H	I	E	Y	L	P	E	V	D	T	G	K	V	K	S	G	E	N	E	L	R	V	V	E	N	R	L	120	

QY	121	KVGGEPPSKVPDSCGHTVGF	FFGSGFP	PANFDF	FPYGGI	IRPVLI	EFTD	HARILDI	WD	TSES	180		
Db	121	KVGGEPPSKVPDSCGHTVGF	FGSGFP	PANFDF	FPYGGI	IRPVLI	EFTD	HARILDI	WD	TSES	180		
QY	181	EPEKKLGKVKVKI	EVSEEA	VGOEMTI	KLGE	EKKIRISNR	FVEGE	FILENAR	FWSL	EDPY	240		
Db	181	EPEKKLGKVKVKI	EVSEEA	VGOEMTI	KLGE	EKKIRISNR	FVEGE	FILENAR	FWSL	EDPY	240		
QY	241	LYPLKVELEKDEY	TLDIGIR	TISWDE	KRLYL	NGKPV	LKGF	GKH	EFPVL	GQGFY	PLMI	300	
Db	241	LYPLKVELEKDEY	TLDIGIR	TISWDE	KRLYL	NGKPV	LKGF	GKH	EFPVL	GQGFY	PLMI	300	
QY	301	KDFNLLKWINANS	FRTSHYP	PYSEEW	LDLAD	RLGIL	VTDEA	PHVGI	TRYHNP	ETOKI	AD	360	
Db	301	KDFNLLKWINANS	FRTSHYP	PYSEEW	LDLAD	RLGIL	VTDEA	PHVGI	TRYHNP	ETOKI	AD	360	
QY	361	NIRRMIDRHKNH	PSVIMSV	ANEPES	NHPDA	EFGFKAL	YETAN	EMDRTR	PVMVSM	MDAP	420		
Db	361	NIRRMIDRHKNH	PSVIMSV	ANEPES	NHPDA	EFGFKAL	YETAN	EMDRTR	PVMVSM	MDAP	420		
QY	421	DERTRDVALKY	FDIVCV	NRYYG	WYIYOG	RIE	BGL	OLEKDI	EEL	YARRK	KPIFVTE	FGAD	480
Db	421	DERTRDVALKY	FDIVCV	NRYYG	WYIYOG	RIE	BGL	OLEKDI	EEL	YARRK	KPIFVTE	FGAD	480
QY	481	AIAGIHYDP	POMFSE	EYQAE	LVEK	TRILL	KDYI	IGTH	VMAF	ADFKT	PQNVRR	PILNHK	540
Db	481	AIAGIHYDP	POMFSE	EYQAE	LVEK	TRILL	KDYI	IGTH	VMAF	ADFKT	PQNVRR	PILNHK	540
QY	541	GVFTRDRQPKL	VAAHVLR	RLMSEV	563								
Db	541	GVFTRDRQPKL	VAAHVLR	RLMSEV	563								

```

RESULT 4
US-10-757-093-20
; Sequence 20, Application US/10757093
; Publication No. US20050153448A1
; GENERAL INFORMATION:
;   APPLICANT: CAMBIA
;   TITLE OF INVENTION: Fungal beta-glucuronidase genes and gene products
;   FILE REFERENCE: 415
;   CURRENT APPLICATION NUMBER: US/10/757,093
;   CURRENT FILING DATE: 2004-01-14
;   NUMBER OF SEQ ID NOS: 37
;   SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
;   LENGTH: 563
;   TYPE: PRT
;   ORGANISM: Thermotoga maritima
US-10-757-093-20

```

[illegible]

Db	241	LYPLKVELKEDEYTLDIGIRTISWDEKRLYLNGKRPVFLKGFGKHEEPVLGGGTYPYBLMI	300
QY	301	KDENLLKWINANSFRTSHYPYSEEWLDDLADRLGILVIDEAPHVGITRHYHNPETQKIAED	360
Db	301	KDENLLKWINANSFRTSHYPYSEEWLDDLADRLGILVIDEAPHVGITRHYHNPETQKIAED	360
QY	361	NIRRMIDRHKHPSVIMSVANEESENHPDAEGFEKALYETANEMDRTRPVVMVSMNDAP	420
Db	361	NIRRMIDRHKHPSVIMSVANEESENHPDAEGFEKALYETANEMDRTRPVVMVSMNDAP	420
QY	421	DETRDVALKYFDIVCVNRYYGWYIYQGRIEEGIQALEKDIIEELYARHKRPIFVTEEGAD	480
Db	421	DETRDVALKYFDIVCVNRYYGWYIYQGRIEEGIQALEKDIIEELYARHKRPIFVTEEGAD	480
QY	481	AIAGIHYPDPQPMSESEYQAEIIVEXTIRLLKKDYIIGTHVWAFADFKTPQNVRRLPINHK	540
Db	481	AIAGIHYPDPQPMSESEYQAEIIVEXTIRLLKKDYIIGTHVWAFADFKTPQNVRRLPINHK	540
QY	541	GVFTRDRQPKLVAVHLRLMSEV	563
Db	541	GVFTRDRQPKLVAVHLRLMSEV	563

```

RESULT 5
US-10-757-093-19
; Sequence 19, Application US/10757093
; Publication No. US20050153448A1
; GENERAL INFORMATION:
; APPLICANT: CAMBIA
; TITLE OF INVENTION: Fungal beta-glucuronidase genes and gene products
; FILE REFERENCE: 415
; CURRENT APPLICATION NUMBER: US/10/757, 093
; CURRENT FILING DATE: 2004-01-14
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 19
; LENGTH: 570
; TYPE: PR1
; ORGANISM: Sulfolobus solfataricus
US-10-757-093-19

```

Query Match	33.7%	Score 1011	DB 18	Length 570
Best Local Similarity	39.2%	Pred. No. 2.2e-75		
Matches 230	Conservative 92	Mismatches 191	Indels 74	Gaps 14
Qy	15	INGVMNLEVTSSKDRP-----IAVPGSWNEQYODLCYEEGPFYKTFYVVPX	60	
Db	11	LQGFWMFKIDNENTGEENGWYKGLSESDIITYVPASWNEQNPBKMDQFSGIAMYQKDLFVSN	70	
Qy	61	XLSEQHIRLFAAVNTDCEVFLNGEKVGENHIEYLPFEVDYTGKVKSGENELRVVENRL	120	
Db	71	DNGNRKAMWVFEGAGYITKLMINGEYGTGHSFTQFKFPIKLKV---NEFNKIV---V	123	
Qy	121	KVGGFSPSKVPDSTHTVGFSGSPRAN-----FDFFPYGIIIRPVLIEFTDHARILDIN	174	
Db	124	KIDNTPSPY-----NLPPARDLNNAAFDFFNYSIGIHRPVYIEFVDECHVEDIT	171	
Qy	175	VDTSESEPEKKLGKVKKIEVSEBAVGQEMTIKLGEEBKIK---RTSNRFVEGEFILEN	230	
Db	172	VYT-----KSYGHLKVEI-LSECNQRFSLRFKLVDKEGRVILNESSNEVFEKD--VNN	222	
Qy	231	ARFWSLEDPLIYPLKVELE---KDEYTLDIGIRTISWDEKRLYLNGKRVPLKGFQKHE	285	
Db	223	VIPWSPDNPIYTLIVEMVYVGNLKDYSYERIGFRDVEVKDKIYILNGKPIFLKGFGRHE	282	
Qy	286	EFPVLGGGTFRPLMIKDFNLKWINANSFRTSHYPYSEEWLADLADRLGILVIDEAP--HV	343	
Db	283	DFPILGKFTYGAVLVRDFYLMRKIGANSFRTSHYPYSNEHLADLADEMGFLVILEPPLCY	342	
Qy	344	GITRHYHNPETQKI-----AEDNIRRMIDRHKNHPSVIMWSVANEPSNHPDAEGF	394	
Db	343	NISRVMSQEIATKMGDVKKFEKVRDITIKEMIRQHKNRSVIMYSVMNEPSSDIREVAEF	402	


```
QY      395 FKALYETANEMDRTRPVVMSMDADBERTDVALKCFDVCVNRYYGWIYQGRIEGL 454
      :      :      :      :      :      :      :      :      :      :
Db      403 IRREVELFKSLDSRPVTFAS-----HRSVRDLALEYVDVISLNYHGWYTEMGDIDSGV 457

QY      455 QALEKDIEELYARH-RKPIFVTEFGADAIAGIHDPQMFSEEYQAEVLEKTIKLLKXD 513
      :      :      :      :      :      :      :      :      :      :
Db      458 KVAIELEIEIHKKPEKPIITTEFGADAIYGLHSDPPQMSEEEYQSEMIRKIYIEALREKD 517

QY      514 YIIGTHWAFADFKTPQNVRRPILNKGVFTRDRQPKLVAHVLRRLW 560
      :      :      :      :      :      :      :      :      :      :
Db      518 YIVGFHIWNFADFRTQPNSRTIILNRKGIPTDRQPKLAIAKVEELF 564
```

RESULT 6

```
US-10-120-145-2
; Sequence 2, Application US/10120145
; Publication No. US20030157684A1
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Kilian, Andrzej
; APPLICANT: Keesee, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/10/120,145
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/149,727
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,263
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-120-145-2
```

```
Query Match      31.0%; Score 929; DB 14; Length 602;
Best Local Similarity 35.6%; Pred. No. 1.7e-68;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY      1 MWRPQRNKKRFFILINGVWNLEV-----TSKDRPIAVPGSWNE--QYQDLC 44
      :      :      :      :      :      :      :      :      :      :
Db      1 MLYPINTETRGVFDLNGVWNFKLDYKGKLEEKWYESKLTDTISMAVPSSYNDIGVTKEIR 60

QY      45 YEEGPFTYKTFYVPKLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPFEVDVTGK 104
      :      :      :      :      :      :      :      :      :      :
Db      61 NHIGYVWEREFVTPAYLKDQRIVLRFSGATHKAIYVYNGELVVEHKGGLPFEEAIIINS 120

QY      105 VKSGENELRVVENRLKVGFPSPKVPDSGHTVGFPGS-----FPPANFDFFPYG 154
      :      :      :      :      :      :      :      :      :      :
Db      121 LRDGMNRVTVAVDNIL-----DSTLPGVGLYSEHHEGLGKVI RNKPNFDFFNYA 170

QY      155 GIIRPVLIETDHARILDIWDTSESEPEKKLGKVKYKIEVSEAVGQEMTIKLG--EEE 212
      :      :      :      :      :      :      :      :      :      :
Db      171 GLHRPVKIYTPFTYVEDISVVTDFNGP--TGTVTYTVDFQ---GKAETVKVSVDEE 223

QY      213 KKIIRTSNRFVEGEFILENARFWSLDEPYLYPLKVELEKDEYTLDI-----GIRTI 267
      :      :      :      :      :      :      :      :      :      :
Db      224 GKVAVSTEGLSGNVEIPNVILMEPLNTLYQIKVELVNDGLTIDVYEEPPGVRTVEVNDG 283

QY      268 RLYNGKPVFLKFGKHEEPVLGQGTFFPLMIKDFNLKWINANSFRTSHYPYSEEWLD 327
      :      :      :      :      :      :      :      :      :      :
Db      284 KFLINNKPFYFKGFGKHEDTPINGRGFNEASNVMDFNILKWIIGANSFRTAHYPSSEELMR 343

QY      328 LADRLGILVIDEAPHVGITRYHYN-----PETQKIA-----EDNIRRMIDR 368
      :      :      :      :      :      :      :      :      :      :
Db      344 LADREGLVVIDETPAVG--HLNFMAATTGLGEGSERVSTWEKIRTFEHHQDVLRELVS 400

QY      369 HKNHPSVIMWSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVMSMDADBERTRDV 427
      :      :      :      :      :      :      :      :      :      :
Db      401 DKNHPSVVMWSIANEAATEEBGAYEYFKPLVELTKELDPQKRPTVIVLFVMAATPE--TDK 458
```

```
QY      428 ALKVFDIVCNRRYYGWIYQGRIEGLQALEKDIEELYARH-RKPIFVTEFGADAI 483
      :      :      :      :      :      :      :      :      :      :
Db      459 VAEIIDLVIALNRNNGWYFDGGBLEAAKVHLR---QEFHAMNKRCPGKPIMITEXGADTVA 515

QY      484 GIHYDPQMFSEEYQAEVLEKTIKLLKQDYIIGTHWAFADFKTPQNVRRPILNKGVF 543
      :      :      :      :      :      :      :      :      :      :
Db      516 GFHDIDPVMFEEYQVEYYQANHVVDFEFENFVEGQAWNFADFATSQGVMRVQGNKKGVF 575

QY      544 TRDRQPKLVAHVLRRLMSEV 563
      :      :      :      :      :      :      :      :      :      :
Db      576 TRDRKPKLAHVFERERTNI 595
```

RESULT 7

```
US-10-364-649-2
; Sequence 2, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-364-649-2
```

```
Query Match      31.0%; Score 929; DB 15; Length 602;
Best Local Similarity 35.6%; Pred. No. 1.7e-68;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY      1 MWRPQRNKKRFFILINGVWNLEV-----TSKDRPIAVPGSWNE--QYQDLC 44
      :      :      :      :      :      :      :      :      :      :
Db      1 MLYPINTETRGVFDLNGVWNFKLDYKGKLEEKWYESKLTDTISMAVPSSYNDIGVTKEIR 60

QY      45 YEEGPFTYKTFYVPKLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPFEVDVTGK 104
      :      :      :      :      :      :      :      :      :      :
Db      61 NHIGYVWEREFVTPAYLKDQRIVLRFSGATHKAIYVYNGELVVEHKGGLPFEEAIIINS 120

QY      105 VKSGENELRVVENRLKVGFPSPKVPDSGHTVGFPGS-----FPPANFDFFPYG 154
      :      :      :      :      :      :      :      :      :      :
Db      121 LRDGMNRVTVAVDNIL-----DSTLPGVGLYSEHHEGLGKVI RNKPNFDFFNYA 170

QY      155 GIIRPVLIETDHARILDIWDTSESEPEKKLGKVKYKIEVSEAVGQEMTIKLG--EEE 212
      :      :      :      :      :      :      :      :      :      :
Db      171 GLHRPVKIYTPFTYVEDISVVTDFNGP--TGTVTYTVDFQ---GKAETVKVSVDEE 223

QY      213 KKIIRTSNRFVEGEFILENARFWSLDEPYLYPLKVELEKDEYTLDI-----GIRTI 267
      :      :      :      :      :      :      :      :      :      :
Db      224 GKVAVSTEGLSGNVEIPNVILMEPLNTLYQIKVELVNDGLTIDVYEEPPGVRTVEVNDG 283

QY      268 RLYNGKPVFLKFGKHEEPVLGQGTFFPLMIKDFNLKWINANSFRTSHYPYSEEWLD 327
      :      :      :      :      :      :      :      :      :      :
Db      284 KFLINNKPFYFKGFGKHEDTPINGRGFNEASNVMDFNILKWIIGANSFRTAHYPSSEELMR 343

QY      328 LADRLGILVIDEAPHVGITRYHYN-----PETQKIA-----EDNIRRMIDR 368
      :      :      :      :      :      :      :      :      :      :
Db      344 LADREGLVVIDETPAVG--HLNFMAATTGLGEGSERVSTWEKIRTFEHHQDVLRELVS 400

QY      369 HKNHPSVIMWSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVMSMDADBERTRDV 427
      :      :      :      :      :      :      :      :      :      :
Db      401 DKNHPSVVMWSIANEAATEEBGAYEYFKPLVELTKELDPQKRPTVIVLFVMAATPE--TDK 458
```


Oy	428	ALKKFDIVCNRRYYGWIYQGRIEEGIALEKDIEELVARRH---	KPIFTEFGADALA	483
	:	: : : :	: : : : :	
Dd	459	VAEIIDIVALNRYNGWYFDGGDLAAKVHLR--QEFTAMNRKCPGPKIMITEYGADTVA		515
Oy	484	GIHDPQPOMSEEQOAEVLVEKTIRLLKKDYIICTHWAFADFKTPONVRPILNHKGVF		543
		: :	: : : :	
Dd	516	GFHIDPVMFTEEYOVEYYQANHVFDEFENFVGEOANNFADFATISQGYMRVOGNKKGVF		575
Oy	544	TRDRQPKLVAAHVLRIMSEV		563
	: :	:		
Dd	576	TRDRKPKLAAHVFRERWTNI		595

```

RESULT 8
US-10-364-649-8
; Sequence 8, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-364-649-8

```

```

Query Match      31.0%; Score 929; DB 15; Length 602;
Best Local Similarity 35.6%; Pred. No. 1.7e-68;
Matches 221; Conservative 93; Mismatches 224; Indels 82; Gaps 15;

QY          1 MVRPQRNKKRFLILINGVWNLV-----TSKDRPIAVPGSWNE--QYODLC 44
           | : | : ||||| :: : : : : : : : : : : : : : : : : : :
DB          1 MLYPINTEIRGVFDLNGVMNFKLIDYGKGLEEKMYESKLTDTTISMAVPSSYNDICVTKEIR 60
QY          45 YEEGPFTYKTTFEVPVKLSQKHRLYFAAVNTDCEVLNGEKYGENHIEYLPREVDVTGK 104
           | : | : ||||| : : : : : : : : : : : : : : : : : : :
DB          61 NHIGVWYEREFTVPAYLKDQRIVLRFGSATHRAIVYNGELVEHKGFLEPAEAINNS 120
QY          105 VKSGENELRVVENRLKVGGFSPKVPDSGTHTVGFFGS-----FPANFDFFPYG 154
           :: : : : : : : : : : : : : : : : : : : : : : : : :
DB          121 LRDMNRVTVAVDNIL-----DDSTLPGLYSERHEEGSKVIRNKPNDFEFNYA 170
QY          155 GIIRPVLIETDHARILDIMVDTSESEPEKKLGKVKVIEVSEAVGOEMTIKG---EEE 212
           | : | | | : : : : : : : : : : : : : : : : : : : : :
DB          171 GLHRPVKIYTTPFTYVEDISVTTDFNGP---TGTVTYTVDFQ----GKAETVKSVDDEE 223
QY          213 KIRTSNRFVEGEFILENARFWSLEDPLYPLKVELEKDEYTLDI-----GIRTISWDEK 267
           | : | : : : : : : : : : : : : : : : : : : : : : : :
DB          224 GKVVASTEGLSGNVEIPNVILMEPLNTYLQIKVELVNDGLTIDVYEEPFGRTVEVNDG 283
QY          268 RLYLNGKRPVFLKFGFKGHEEPVYLGQGTFFPLMIKDNLTKWINANSFRTSHYPYSEEWLD 327
           : : | : : : : : : : : : : : : : : : : : : : : : : :
DB          284 KPLINNKPFEYFGKGHEDTPINGRGFNASNVMDNILKWIGANSFRTAHYPYSEELMR 343
QY          328 LADRLGILVIDEAPHVGITRYHYN-----PETOKIA-----EDNIRMIDR 368
           ||| : : : : : : : : : : : : : : : : : : : : : : : :
DB          344 LADREGLVVIDETPAVG--HLNFMATTGLGGSERSVSTWEKIRTFEHHQDLRELVS 400
QY          369 HKNHPSVIMSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVMVSMMDABDERTRDV 427
           ||||| : : : : : : : : : : : : : : : : : : : : : : :
DB          401 DKNHPSVVMMSIANEAATEEGAYEYFKPLVELTKELDPOKRPVTI VLFVMAYPE--TDK 458
QY          428 ALKYFDIVCVNRRYYGIYOGRIEEGLQALEKDI EELYARHR---KPIFVTEFGADAIA 483

```

Db 459 VAEIIDLVIATLNRTNGWYFDGGDLEAAKVHLR---QEFHAMNKRCPGKPFIMITEYGADTVA 515

QY 484 GIHYDPQPMFSEYQABLVEKTRILLKKDYITGHWAFADFKTPONVRRPILNHKGVF 543

Db 516 GFHDIDPVMFTSEYQVEYYQANHVFDEFENFVGEOAWNFAFATISQGMRVQGNKKGVF 575

QY 544 TRDRQPKLVAAHVLRLMSEV 563

Db 576 TRDRKPKLAAHVFRERWTNI 595

```

RESULT 9
US-10-364-649-15
; Sequence 15, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge
; TITLE OF INVENTION: MICROBIAL B-GLUCURONI
; TITLE OF INVENTION: PRODUCTS, AND USES T
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 602
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-364-649-15

```

	Query Match	31.0%;	Score 929;	DB 15;	Length 602;	
	Best Local Similarity	35.6%;	Pred. No. 1.7e-68;			
	Matches 221;	Conservative 93;	Mismatches 224;	Indels 82;	Gaps 15;	
QY	1 MVRPORNKKRFFILILINGVNWLEV-----TSKDRPIAVPGSWNE--QYODLC	44				
DB	1 MLYPINTETRGVFDLNGVWNPFLDYGKGLEEKWYESKLTDITISMAVPSSYNDIGVTKEIR	60				
QY	45 YEEGPFTYKTTFYVPKXLSQHIRLYFAAVNTDCEVFPLNGEKEVGEMHIEYLPEVDVTGK	104				
DB	61 NHIGYWWEREFTVPAYLKDQRIVLRFGSATPKAIVYVNGELVVEHKGFAPFEABINNS	120				
QY	105 VKSGENELRVVENRLKVGGFPPSKVPDSGTHTVGFFGS-----FPANPDFFPYG	154				
DB	121 LRDMNRVVAVDNII-----DDSTLPVGLYSERHEGLKVIARNKNPDFFNYA	170				
QY	155 GIIRPVLIEFTDHARILDIMVDTSESEPEKKLGKVKKIEVSEAVGOEWTKLG--EEE	212				
DB	171 GLHRPVKIYTTPFTYVEDISVTTDENGPR---TGTVTYTVDFQ---GKAETVKSVDDEE	223				
QY	213 KKIRTSNRFVEGEFILLENARFWSLEDPLYLPLKVELKEDEYTLDI----GIRTISWDEK	267				
DB	224 GKVVASTEGTSGNVETIPNVILLMEPLNTLYIQIKVELVNDGITIDYEEEPFGVRTVEVNDG	283				
QY	268 RLYLNGKPVFLKGFGRHEEPVLGGTFYFLMIKDPNLKMINANSFRTSHTPYSEBWL	327				
DB	284 KFLINNKPFFYFKGFGRKHEDTPIINGRGFNEASNMDFNIKMGANSFRTHAYPYSEBLMR	343				
QY	328 LADRLGILVIDEAPHVGITRIHYN-----PETOKIA-----EDNIRRMIDR	368				
DB	344 LADREGLVVIDETPAVG---HLNFMATGTGLGEGSERVSITWEKIRTFPEHHODVLRVSR	400				
QY	369 HKNHPSVIMWSVANEPESNHPDAEGFFKALYETANEMD-RTRPVVMVMSMDAPDERTRDV	427				
DB	401 DKNHPSVVMWSIANEAATEEBGALEYEFKPLVELTKELDPOKRPTVITLVFWMATPE--TDK	458				
QY	428 ALKYFDIVCVNRYRGWIYOGRIEEGLQALEKDIEELYARHR---KPIFVTEFGADAIA	483				

Db 459 VAEILDVIALNRRYNGWYFDGGDLEAAKVHLR--QEFHANMKRCPGKPIMITTEYGADTVA 515

Qy 484 GIHYDPQMFSESEYQALVEKTRILLKKDYIIGTHVAFADEKTPQNVRRPILNHKGVF 543

Db 516 GFHDIDPMFTSEYQVEYYQANHVFDEFENFVGEQANMFADPATSQGVNRVQGNKKGVF 575

Qy 544 TRDRQPKLVAAHLRLRMSEV 563

Db 576 TRDRKPKLAAHVFRRWTNI 595

RESULT 10

```

US-10-757-093-23
/ Sequence 23, Application US/10757093
/ Publication No. US20050153448A1
/ GENERAL INFORMATION:
/ APPLICANT: CAMEBA
/ TITLE OF INVENTION: Fungal beta-glucuronidase genes and gene products
/ FILE REFERENCE: 415
/ CURRENT APPLICATION NUMBER: US/10/757,093
/ CURRENT FILING DATE: 2004-01-14
/ NUMBER OF SEQ ID NOS: 37
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 23
/ LENGTH: 602
/ TYPE: PRT
/ ORGANISM: Staphylococcus sp.
US-10-757-093-23

```

Query Match	31.0%;	Score 929;	DB 18;	Length 602;
Best Local Similarity	35.6%;	Pred. No. 1.7e-68;		
Matches 221; Conservative	93;	Mismatches 224;	Indels 82;	Gaps 15;

Qy	1	MWRPQNRKKRFFILLNGVWNLEV-----TSKDRPIAVGSMWE--QYQDLC	44
Db	1	MLYPINTETRGVFDLNGVWNFKLDYGKGLEEKWYESKLTDTISMAVPSSYNDIGVTKEIR	60
Qy	45	YEEGDFYKTFYVVPKLSQKHIRLYFAAVNTDCEVFLNGEKVGENHIEYLPEFVDVTGK	104
Db	61	NHIGYVWEREFYTPAYLKDQRIVLRFGSATHKAIYVYNGELVVEHKGGLPFEAEINNS	120
Qy	105	VKSGENELRVVVENRLKVGGFSPSKVPDSCGHTVGFSGS-----FPANPDFFPYG	154
Db	121	LRDGNMRVTVAVDNL-----DDSTLPVGLYSEHHEGLGKVI RNKNPNDFFNVA	170
Qy	155	GIIRPVLIETDHDARILDIWDTSESEPEKKLGKVKVKEVSEAVQEWITKLG--EEE	212
Db	171	GLHRPVKIYTPFTYVEDISVTDFNGP---TGTVTYTVDFO---GKAETVKVSVDEE	223
Qy	213	KKIRPSNRFEVGEFILENARFWSLEDPYLPKVELEKDEYTLDI-----GIRTISWDEK	267
Db	224	GKVAVSTEGLSGNVEIPNVILMEPLNTLYQIKVELVNDGLTIDVYEPRGVRTVEVNDG	283
Qy	268	RLYLNGKRPVFLKGFQKHEEPVLGGCTFYPLMIKD FNLLKMINANSFRTSHYPYSEEWLD	327
Db	284	KFLINNKPFYFKGFGKHEDTPINGRGFNEASNVMDFNILKMI GANSFRTAHYPYSEELMR	343
Qy	328	LADRLGILVIDEAPHVGITRYHYN-----PETOKIA-----EDNIRRMIDR	368
Db	344	LADREGLVVIDETPAVG--HLPNFMATTGLGEGSERVSTWEKIRTFEHHQDVLRELVS	400
Qy	369	HKNHPSVIMSVVANEPSNNHPDAEGFFALYETANEMD-RTRPVVMVSMMDAPDERTDV	427
Db	401	DKNHSVVMWSIANEAATEEGAYEYFKPLVELTKELDPQKRPVTIVLFVMAATPE--TDK	458
Qy	428	ALKYFDIVCVNRYGYWYIYQRIEELGQALEKDIIEELYARHR---KPIFVTEFGADAIA	483
Db	459	VAELIDVIALNRYNGWYFDGGDLEAAKVHLR---QEFHAMNKRCPGKPIMITTEYGADTVA	515
Qy	484	GIHYDPEQMFSESYQAEIVKERTIRLLKKDYIIGTHVMAFADEKTPQNVARRPILNHKGVF	543
Db	516	GFHDIDVPMFTEEYQVEYYQANHVFDEEENFVGEQANMFADFATSGVNRVOGNKKGVF	575

```
Qy      544 TRDRQPLVAHVLRRLMSEV 563
        |||:|||||:|:
Db      576 TRDRKPKLAHVFRERWTNI 595
```

RESULT 11

```

US-10-120-145-4
; Sequence 4, Application US/10120145
; Publication No. US20030157684A1
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Kilian, Andrzej
; APPLICANT: Keese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/10/120,145
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/149,727
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,263
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 618
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-120-145-4

```

Query Match	31.0%;	Score 929;	DB 14;	Length 618;
Best Local Similarity	35.6%;	Pred. No. 1.7e-68;		
Matches 221; Conservative	93;	Mismatches 224;	Indels 82;	Gaps 15;

QY	1	MVRPQNRKKR	FILLINGWN	LEV-----	TSKDRPIAVPGS	WNE--QY	0DLC	44
		: :	: :	: :	: :	: :	: :	
Db	17	MLYPINTETRG	VFDLNGWN	NFKLDY	GKGLEEKW	YESK	LDTISMAVPS	SNDIGVTKEIR 76
QY	45	YEEGPFTYKTT	FYVPKXLSQ	KHIRLYFAA	VNTDC	CEVFLNGEK	VGENHIEYL	PFEVDVTGK 104
		: :	: :	: :	: :	: :	: :	
Db	77	NHIGVWYERE	FETPAVYLK	DQRIVLRF	GSATHKAI	VYVNGEL	VVEHKGGL	PFEAEIINNS 136
QY	105	VKSGENELRV	VVENRLKV	GGFSPSKV	PD	SGTHTV	GFPGS-----	FRPANFDFPYG 154
		: :	: :	: :	: :	: :	: :	
Db	137	LRDGNRVTVA	VDNL-----	DDSTLPVGL	YSERHBEGL	GKVTIR	KNPNDF	FENYA 186
QY	155	GIRPVLIEFT	DHARLIDI	WVDTSE	SEPEKKLG	KVVKYIEV	SEEA	VQGE
		: :	: :	: :	: :	: :	: :	
Db	187	GLHREPKI	YTPFTY	VEDISV	VTDFNGP--	TGTVYTV	DFQ----	GKAETVKVSVDEE 239
QY	213	KKIRTSNR	FVEGEFI	ENARFWS	LEDPYLP	LKVELEK	DEYTLDI----	GIRTISWDEK 267
		: :	: :	: :	: :	: :	: :	
Db	240	GKVVA	STEGLSGN	VEIPN	VILMEPL	NTLYLOIK	VELVND	GLTIDVYEE
QY	268	RLYLNGK	PVFLKFG	KEEFPVL	GGGTFF	PLMIK	DNL	LKMINANS
		: :	: :	: :	: :	: :	: :	
Db	300	KFLIN	KKPFFY	KFGKHED	TPINGR	GFNEAS	NVMD	FNILK
QY	328	LADRLGIL	VIDEAPHV	GITRYHN-----	PETOKIA----	EDNIR	RMIDR	368
		: :	: :	: :	: :	: :	: :	
Db	360	LADREGL	VIDETPA	VGV--H	LNEMATT	GLGEGSER	VS	TWEKIRTFE
QY	369	HKNHDS	VIMSV	ANES	NHPDAEG	FPKALY	ETAN	END--RTRP
		: :	: :	: :	: :	: :	: :	
Db	417	DKNHS	VVMWS	IANEA	TEEGAYE	FKPLVEL	TKELD	PQKR
QY	428	ALKYED	IVCVN	RYGMYI	YOGRIE	EGLQALEK	DI	EELV
		: :	: :	: :	: :	: :	: :	
Db	475	VAELID	VIALN	RYNGWY	FDGDL	EAAKVHLR--	QEFHAM	NKRC
QY	484	GIHYDP	QOMF	SEEQAE	LVEK	IRLLLK	DYIIG	THVAFAD
		: :	: :	: :	: :	: :	: :	
Db	532	GFHDID	PVMT	TEEYOE	YVYOA	NHVVE	DEFEN	FVGE
		: :	: :	: :	: :	: :	: :	

QY 544 TRDRQPLVAHLRRLWSEV 563
Db 592 TRDRKPKLAHVFRRERWTNI 611

RESULT 12

US-10-120-145-8
; Sequence 8, Application US/10120145
; Publication No. US20030157684A1
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Kilian, Andrzej
; APPLICANT: Keese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/10/120,145
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/149,727
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,263
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 607
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-120-145-8

Query Match 30.8%; Score 925; DB 14; Length 607;
Best Local Similarity 35.7%; Pred. No. 3.6e-68;
Matches 220; Conservative 92; Mismatches 223; Indels 82; Gaps 15;

QY 4 PQRNKKRFFILINGVWNLV-----TSKDRPIAVPGSWE--QYQDLCYEE 47
Db 9 PINTETRGVFDLNGVWNFKLDYKGLEBKWYESKLTDTISMAVPSYNDIGVTKELRNHI 68
QY 48 GPFTYKTTFFVVPKXLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPFEVDVTGKYKS 107
Db 69 GYWYEREFVPAVLKDQRIVLRFGSATHKAIYVNGELVVEHKGFLPFEAEINNSLRD 128
QY 108 GENELRVVENRLKVGGFPSKVPDSCGTHTVGFFGS-----PPANFDFFPYGGII 157
Db 129 GMRVTVAVDNIL-----DDSTLPVGLYSEHHEGLGKVI RNKPNFDFFNYAGLH 178
QY 158 RPYLIEFTDHARILDIWVDTSESEPEKKLGKVKKIEVSEAVGOEMTIKLG--EEEKKI 215
Db 179 RPVKIYTTPTTYVEDISVTDENGPR--TGTVTYTVDFQ---GKAETVKSVDDEGKV 231
QY 216 RTSNRFEVEGEFILENARFWSLEDPYLPKLEKDEYTLDI-----GIRTISWDEKRLY 270
Db 232 VASTEGLSGNVEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYEBPFGVRTVEVNDGKFL 291
QY 271 LNKRPVFLKGFKGHEEPVLGQGTFFPLMIKDFNLKMINANSFRTSHYPYSEEWLAD 330
Db 292 INNKPFYFKGFGKHEDTPINGRGFNEASNVMDFNLKMWIGANSFRTAHYPYSEELMLAD 351
QY 331 RLGLIVIDEAPHVIGITRYHYN-----PETOKIA-----EDNIRMIDRHKH 371
Db 352 REGLVVIDETPAVG--HLNFMAITGLGEGSERVSTWEKIRTFEHHQDVLRELVS RDKN 408
QY 372 HPSVIMSVANPEPSNHPDAEGFFKALYETANEMD--RTRPVVMVSMADAPDERTRVALK 430
Db 409 HPSVVMWSIANEAATEEGAYEYFKPLVELYTKELDPQKRPTIIVLFWMATPE--TDKVAE 466
QY 431 YFDIVCVNRYGWIYIOGRIEGLQALEKDI EELVYARH---KPIFVTEFGADAIAGIH 486
Db 467 LIDVIALNRNGWYFDGDL EAAKVHLR---QEFHAMNKRCPGKPIMITTEYGADTVAGFH 523
QY 487 YDPPQMFSEEYQAEI LVEKTI RLLKKDYIIGTHVWAFADFKTPQNVRRPILNHKGVFTRD 546
Db 524 DIDVPMFTEEYQVEYIYQANHVFDEFEENFVGEQAMNFPADFATSQGVMRVQGNKKGVFTRD 583

Db 524 DIDPVNFTEEYQVEYIYQANHVFDEFEENFVGEQAMNFPADFATSQGVMRVQGNKKGVFTRD 583
QY 547 RQPKLVAHLRRLWSEV 563
Db 584 RKPPLAHLVFRERWTNI 600

RESULT 13

US-10-364-649-18
; Sequence 18, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 607
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-10-364-649-18

Query Match 30.8%; Score 925; DB 15; Length 607;
Best Local Similarity 35.7%; Pred. No. 3.6e-68;
Matches 220; Conservative 92; Mismatches 223; Indels 82; Gaps 15;

QY 4 PQRNKKRFFILINGVWNLV-----TSKDRPIAVPGSWE--QYQDLCYEE 47
Db 9 PINTETRGVFDLNGVWNFKLDYKGLEBKWYESKLTDTISMAVPSYNDIGVTKELRNHI 68
QY 48 GPFTYKTTFFVVPKXLSQKHIRLYFAAVNTDCEVFLNGEKGVENHIEYLPFEVDVTGKYKS 107
Db 69 GYWYEREFVPAVLKDQRIVLRFGSATHKAIYVNGELVVEHKGFLPFEAEINNSLRD 128
QY 108 GENELRVVENRLKVGGFPSKVPDSCGTHTVGFFGS-----PPANFDFFPYGGII 157
Db 129 GMRVTVAVDNIL-----DDSTLPVGLYSEHHEGLGKVI RNKPNFDFFNYAGLH 178
QY 158 RPYLIEFTDHARILDIWVDTSESEPEKKLGKVKKIEVSEAVGOEMTIKLG--EEEKKI 215
Db 179 RPVKIYTTPTTYVEDISVTDENGPR--TGTVTYTVDFQ---GKAETVKSVDDEGKV 231
QY 216 RTSNRFEVEGEFILENARFWSLEDPYLPKLEKDEYTLDI-----GIRTISWDEKRLY 270
Db 232 VASTEGLSGNVEIPNVILMEPLNTYLYQIKVELVNDGLTIDVYEBPFGVRTVEVNDGKFL 291
QY 271 LNKRPVFLKGFKGHEEPVLGQGTFFPLMIKDFNLKMINANSFRTSHYPYSEEWLAD 330
Db 292 INNKPFYFKGFGKHEDTPINGRGFNEASNVMDFNLKMWIGANSFRTAHYPYSEELMLAD 351
QY 331 RLGLIVIDEAPHVIGITRYHYN-----PETOKIA-----EDNIRMIDRHKH 371
Db 352 REGLVVIDETPAVG--HLNFMAITGLGEGSERVSTWEKIRTFEHHQDVLRELVS RDKN 408
QY 372 HPSVIMSVANPEPSNHPDAEGFFKALYETANEMD--RTRPVVMVSMADAPDERTRVALK 430
Db 409 HPSVVMWSIANEAATEEGAYEYFKPLVELYTKELDPQKRPTIIVLFWMATPE--TDKVAE 466
QY 431 YFDIVCVNRYGWIYIOGRIEGLQALEKDI EELVYARH---KPIFVTEFGADAIAGIH 486
Db 467 LIDVIALNRNGWYFDGDL EAAKVHLR---QEFHAMNKRCPGKPIMITTEYGADTVAGFH 523
QY 487 YDPPQMFSEEYQAEI LVEKTI RLLKKDYIIGTHVWAFADFKTPQNVRRPILNHKGVFTRD 546
Db 524 DIDVPMFTEEYQVEYIYQANHVFDEFEENFVGEQAMNFPADFATSQGVMRVQGNKKGVFTRD 583

Job time : 170 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2005, 19:13:30 ; Search time 7488 Seconds
(without alignments)
10929.602 Million cell updates/sec

Title: US-09-936-759-14
Perfect score: 1689
Sequence: 1 atggttaagaccgcaacgaaa.....gaagactgtgagtgaggtc 1689

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_rts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1688	99.9	1689	6 AR428149	AR428149 Sequence
2	1688	99.9	12583	1 AE001766	AE001766 Thermotog
3	194	11.5	2153	4 AF012423	AF012423 Felis cat
4	192.4	11.4	2014	4 AF012424	AF012424 Felis cat
5	190.2	11.3	2199	4 AF019759	AF019759 Canis fam
6	174.4	10.3	2155	9 AF084552	AF084552 Chloroceb
7	174	10.3	2128	6 CQ575703	CQ575703 Sequence
8	174	10.3	2182	9 BC014142	BC014142 Homo sapi
9	174	10.3	2321	3 BT010073	BT010073 Drosophil
10	172.4	10.2	1956	6 AR344264	AR344264 Sequence
11	172.4	10.2	1956	6 AX147652	AX147652 Sequence
12	172.4	10.2	2191	6 CQ847932	CQ847932 Sequence
13	172.4	10.2	2191	9 HUMGLCB	M15182 Human beta-
14	166.8	9.9	2391	10 RINGLUCB	Y00717 Rat mRNA fo
15	166.8	9.9	2472	6 AX827491	AX827491 Sequence
16	166.8	9.9	2472	10 RATGICB	M13962 Rat beta-g1
17	159.6	9.4	1947	6 AR344265	AR344265 Sequence
18	159.6	9.4	1947	6 AX147654	AX147654 Sequence
19	159.6	9.4	2456	10 MUSGUS	M28540 Mouse beta-

20	159.6	9.4	2456	10 MUSGUS	M28541 Mouse beta-
21	159.6	9.4	2456	10 MUSGUS	M63836 Mouse beta-
22	159.6	9.4	2492	10 BC071226	BC071226 Mus muscu
23	157.8	9.3	2454	10 MUSGLC	M19279 Mouse beta-
24	156.4	9.3	2455	10 MUSGLC	J03047 Mouse beta-
25	154.8	9.2	2081	5 AJ720880	AJ720880 Gallus ga
26	142.2	8.4	23078	6 CQ363720	CQ363720 Sequence
27	142.2	8.4	60265	1 AE017283	Continuation (26 o
28	141.4	8.4	1809	12 AF354047	AF354047 Synthetic
29	141.4	8.4	1887	6 AR210802	AR210802 Sequence
30	141.4	8.4	1887	6 BD074927	BD074927 Microbial
31	141.4	8.4	7018	12 AY452736	AY452736 Reporter
32	141.4	8.4	11846	12 AF354045	AF354045 Binary ve
33	141.4	8.4	11921	12 AF354046	AF354046 Binary ve
34	141.2	8.4	2235	11 BV175016	BV175016 sqm76123
35	139	8.2	10775	1 AE006894	AE006894 Sulfolobu
36	138.6	8.2	2352	3 BT012475	BT012475 Drosophil
37	138.6	8.2	2445	3 AY069329	AY069329 Drosophil
38	130.4	7.7	1888	6 AR428153	AR428153 Sequence
39	127.6	7.6	93590	3 AC084452	AC084452 Caenorhab
40	123.6	7.3	301450	1 AP003185	AP003185 Clostridi
41	122	7.2	3448	1 CPE420784	AJ420784 Clostridi
42	115.6	6.8	2105	1 AF305917	AF305917 Escherich
43	115.6	6.8	300330	1 AP005222	AP005222 Corynebac
44	114	6.7	1814	6 BD185061	BD185061 Nucleic a
45	114	6.7	2103	1 AF305918	AF305918 Escherich

ALIGNMENTS

RESULT 1
AR428149 1689 bp DNA linear PAT 18-DEC-2003
LOCUS Sequence 14 from patent US 6641996.
DEFINITION AR428149
ACCESSION AR428149 GI:40187542
VERSION AR428149.1 GI:40187542
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1689)
AUTHORS Jefferson,R.A. and Mayer,J.E.
TITLE Microbial .beta.-glucuronidase genes, gene products and uses thereof
JOURNAL Patent: US 6641996-A 14 04-NOV-2003;
FEATURES Location/Qualifiers
source 1. 1689
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match	Best Local Similarity	Score	Pred.	DB	Length	Matches	Conservative	Mismatches	Indels	Gaps
1689	100.0%	1688	0	6	1689	1689	0	0	0	0
1		1688	99.9%	6	1689	1689	0	0	0	0
2		1688	99.9%	1	12583	12583	0	0	0	0
3		194	11.5%	4	2153	2153	0	0	0	0
4		192.4	11.4%	4	2014	2014	0	0	0	0
5		190.2	11.3%	4	2199	2199	0	0	0	0
6		174.4	10.3%	9	2155	2155	0	0	0	0
7		174	10.3%	6	2128	2128	0	0	0	0
8		174	10.3%	9	2182	2182	0	0	0	0
9		174	10.3%	3	2321	2321	0	0	0	0
10		172.4	10.2%	6	1956	1956	0	0	0	0
11		172.4	10.2%	6	1956	1956	0	0	0	0
12		172.4	10.2%	6	2191	2191	0	0	0	0
13		172.4	10.2%	9	2191	2191	0	0	0	0
14		166.8	9.9%	10	2391	2391	0	0	0	0
15		166.8	9.9%	6	2472	2472	0	0	0	0
16		166.8	9.9%	10	2472	2472	0	0	0	0
17		159.6	9.4%	6	1947	1947	0	0	0	0
18		159.6	9.4%	6	1947	1947	0	0	0	0
19		159.6	9.4%	10	2456	2456	0	0	0	0

```
|||||
Db      241 TTCTCAACGAGAGAAAAGTGGAGAGAAATCACATTAATACCTTCCCTTGAAGTAGAT 300
Qy      301 GTGACGGGGAAAGTGAATCCCGAGAGAACGAACTCAGGGTGTGTGGAACAGATTG 360
Db      301 GTGACGGGGAAAGTGAATCCCGAGAGAACGAACTCAGGGTGTGTGGAACAGATTG 360
Qy      361 AAAGTGGAGGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGATT 420
Db      361 AAAGTGGAGGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGATT 420
Qy      421 GGAAGTTTCCACCTGCAAACTTCGACTTCTTCCCTACGGTGAATCATTAAGCCTGTT 480
Db      421 GGAAGTTTCCACCTGCAAACTTCGACTTCTTCCCTACGGTGAATCATTAAGCCTGTT 480
Qy      481 CTGATAGAGTTCACAGACCACGCGAGATTAATCTGATCTGGTGAACAGAGTAGTCT 540
Db      481 CTGATAGAGTTCACAGACCACGCGAGATTAATCTGATCTGGTGAACAGAGTAGTCT 540
Qy      541 GAAACGGAGAAAGAACTTGAAAAAGTGAAGTAGAAGTCTCAGAAAGACGGTG 600
Db      541 GAAACGGAGAAAGAACTTGAAAAAGTGAAGTAGAAGTCTCAGAAAGACGGTG 600
Qy      601 GACACGAGATGACGATCAAACTTGAGAGAGAAAGAAAGATTAGAATCCACAGA 660
Db      601 GACACGAGATGACGATCAAACTTGAGAGAGAAAGAAAGATTAGAATCCACAGA 660
Qy      661 TTCTGGAAGGGAGTTTCATCTCTCGAAAAACGCCAGTTCTGAGCCTCGAAGATCCATAT 720
Db      661 TTCTGGAAGGGAGTTTCATCTCTCGAAAAACGCCAGTTCTGAGCCTCGAAGATCCATAT 720
Qy      721 CTTTATCCTCTCAAGGTGGAACCTTGAAAAAGACGAGTACACTCTGGAATCGAATCAGA 780
Db      721 CTTTATCCTCTCAAGGTGGAACCTTGAAAAAGACGAGTACACTCTGGAATCGAATCAGA 780
Qy      781 ACGATCAGCTGGAGACGAGAGAGGCTTATCTGAAACGGAACCTGTCTTTTGAAGGC 840
Db      781 ACGATCAGCTGGAGACGAGAGAGGCTTATCTGAAACGGAACCTGTCTTTTGAAGGC 840
Qy      841 TTTGGAAGACGACGAGAAATCCCGTCTGGGGAGGGCACCTTTATCCATTGATGATA 900
Db      841 TTTGGAAGACGACGAGAAATCCCGTCTGGGGAGGGCACCTTTATCCATTGATGATA 900
Qy      901 AAAGACTTCAACCTTCTGAAGTGAATCAACGCGAATTTCTTACAGACCTTCACTATCCT 960
Db      901 AAAGACTTCAACCTTCTGAAGTGAATCAACGCGAATTTCTTACAGACCTTCACTATCCT 960
Qy      961 TACAGTGAAGAGTGGCTGGATCTTGCGACAGACTCGGAATCCTTGATAGACGAAGCC 1020
Db      961 TACAGTGAAGAGTGGCTGGATCTTGCGACAGACTCGGAATCCTTGATAGACGAAGCC 1020
Qy      1021 CCGCAGTTGTATCAACAAGTACCACTAATCCCGAGACTCAGAAAGATAGACGAAGAC 1080
Db      1021 CCGCAGTTGTATCAACAAGTACCACTAATCCCGAGACTCAGAAAGATAGACGAAGAC 1080
Qy      1081 AACATAAGAAAGATGATCGACAGACACAAGAACCAATCCCAAGTGTATCATGTGAGTGTG 1140
Db      1081 AACATAAGAAAGATGATCGACAGACACAAGAACCAATCCCAAGTGTATCATGTGAGTGTG 1140
Qy      1141 GCGAAGCAACGAGTCCAAACCATCGACGCGGAGGTTTCTCAAGCCCTTATGAG 1200
Db      1141 GCGAAGCAACGAGTCCAAACCATCGACGCGGAGGTTTCTCAAGCCCTTATGAG 1200
Qy      1201 ACTGCCAATGAATGATCGAACACGCCCCGTTGTATGTGATGATGATGACGACCA 1260
Db      1201 ACTGCCAATGAATGATCGAACACGCCCCGTTGTATGTGATGATGATGACGACCA 1260
Qy      1261 GACGAGAGAACAGACGCTGCGCTGAAGTACTTGCATCGATCGTGTGTGAACAGGTAC 1320
Db      1261 GACGAGAGAACAGACGCTGCGCTGAAGTACTTGCATCGATCGTGTGTGAACAGGTAC 1320
Qy      1321 TACGGCTGTACATCTATCAGGGAAGATAGAAGAGACTTCAAGCTCTGAAAAAGAC 1380
Db      1321 TACGGCTGTACATCTATCAGGGAAGATAGAAGAGACTTCAAGCTCTGAAAAAGAC 1380
```

```
Db      1321 TACGGCTGTACATCTATCAGGGAAGATAGAAGAGACTTCAAGCTCTGAAAAAGAC 1380
Qy      1381 ATAGAAGAGCTCTATGCAAGGCACAGAAAGCCCATCTTTGTACAGAAATTCCGTGGCAG 1440
Db      1381 ATAGAAGAGCTCTATGCAAGGCACAGAAAGCCCATCTTTGTACAGAAATTCCGTGGCAG 1440
Qy      1441 GCGATAGCTGGCATCCACTACGATCCACCTCAATGTTCTCCGAAGAGTACCAAGCAGAG 1500
Db      1441 GCGATAGCTGGCATCCACTACGATCCACCTCAATGTTCTCCGAAGAGTACCAAGCAGAG 1500
Qy      1501 CTCGTTGAAAAGACGATCAGGCTCCTTTGAAAAAAGACTACATCATCGGAACACACGTG 1560
Db      1501 CTCGTTGAAAAGACGATCAGGCTCCTTTGAAAAAAGACTACATCATCGGAACACACGTG 1560
Qy      1561 TGGGCTTTGACGATTTTAAGACTCCTCAGAAATGTGAGAAGACCCATTCTCAACCAAG 1620
Db      1561 TGGGCTTTGACGATTTTAAGACTCCTCAGAAATGTGAGAAGACCCATTCTCAACCAAG 1620
Qy      1621 GGTGTTTTCACAAGACAGACCAACCAACTCGTTGCTCATGTACTGAGAAGACTGTGG 1680
Db      1621 GGTGTTTTCACAAGACAGACCAACCAACTCGTTGCTCATGTACTGAGAAGACTGTGG 1680
Qy      1681 AGTGAGGTT 1689
Db      1681 AGTGAGGTT 1689
```

```
RESULT 2
AE001766/c      12583 bp      DNA      linear      BCF 04-JUN-2004
LOCUS      AE001766
DEFINITION      Thermotoga maritima MSB8 section 78 of 136 of the complete genome.
ACCESSION      AE001766 AE000512
VERSION      AE001766.1 GI:4981600
KEYWORDS
SOURCE
ORGANISM      Thermotoga maritima MSB8
                Thermotoga maritima MSB8
REFERENCE
AUTHORS      Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
                Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
                McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
                Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
                Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
                Eisen,J.A., White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and
                Fraser,C.M.
                Evidence for lateral gene transfer between Archaea and bacteria
                from genome sequence of Thermotoga maritima
                Nature 399 (6734), 323-329 (1999)
JOURNAL
MEDLINE      99287316
PUBMED      10360571
REFERENCE
AUTHORS      Nelson,K.E., Clayton,R.A., Gill,S.R., Gwinn,M.L., Dodson,R.J.,
                Haft,D.H., Hickey,E.K., Peterson,J.D., Nelson,W.C., Ketchum,K.A.,
                McDonald,L., Utterback,T.R., Malek,J.A., Linher,K.D., Garrett,M.M.,
                Stewart,A.M., Cotton,M.D., Pratt,M.S., Phillips,C.A.,
                Richardson,D., Heidelberg,J., Sutton,G.G., Fleischmann,R.D.,
                White,O., Salzberg,S.L., Smith,H.O., Venter,J.C. and Fraser,C.M.
                Direct Submission
                Submitted (01-JUN-1999) The Institute for Genomic Research, 9712
                Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 12583
    /organism="Thermotoga maritima MSB8"
    /mol_type="genomic DNA"
    /strain="MSB8"
    /db_xref="taxon:243274"
    81..1607
    /locus_tag="TM1058"
    81..1607
    /locus_tag="TM1058"
    /note="similar to GB:AE000666 percent identity: 48.70;
    identified by sequence similarity; putative"
    /codon_start=1
```

/transl_table=11
/product="glutamate synthase-related protein"
/protein_id="AAD36135.1"
/db_xref="GI:4981601"
/translation="MSGICAKCVDGCGTGCETWLASFRGREVLYPGPEDITAGAVKD
YPDVSHNLIGYARGAEGLEGEVPGPDATFTNVDTTTEGMDIKVMKVPITFGA
LGSTEIARKNWDHIAVGAALISGITVVCENAVGVDPDLELDSNGKAKSPELDRIEI
YKRHDGEYGEILIQMNVEDRTLGAAYINKGIETIELKMGQAKSIGEIKVRS
ERALELKRGIYIVLPDELPPEVQRAFKGEIKEFERHSRLGFVSKESFLKEVERLKL
GFKRITLTKGAYSAAVELAMALRYGAABAKVDLITDGA PGGTGMSPMNMENWGIPTFY
LEALTYPAEKLSRGRIVPDIAIAGFSTEDGVLKAIAMGSPYKAVCMGRALMIPA
MVGKNIGEWLKSGLPKTVSKYGTVEELFTVEELRSRGEEVKLPLGAIGVYTF
VQFKTGIQQLMAGARKRILSALSRKDILALTKDAEISGIPYMESYRDEAERILEE
"
complement(1650. .1913)
/locus_tag="TM1059"
complement(1650. .1913)
/locus_tag="TM1059"
/note="similar to SP:P45693 PID:862985 GB:AL009126 percent
identity: 48.70; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="spovs-related protein"
/protein_id="AAD36136.1"
/db_xref="GI:4981602"
/translation="MEVLKVSXSDPNKVGATAGVREHGKAEIQAIAGAVNQAVK
AIAIARGYLAPSGIDLVPVPAFTVEIENEKRTAIKFIVPPKS"
complement(1962. .3272)
/locus_tag="TM1060"
complement(1962. .3272)
/locus_tag="TM1060"
/note="similar to GB:AE00782 percent identity: 49.74;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAD36144.1"
/db_xref="GI:4981610"
/translation="MNISIIEGALAVLINQPFAGAYLTGYFLMGASSFFIGLFGSIP
FLANTLQULTLSFSHRKSRKQIIVPLMTARTSILFAVPAIKHGLLAYLYFYI
QIAGALSAPLMQSWMSDLVPKDMIGSYFGFRNLHGAVQIPAMPLAGAILDSLGENWK
GFGTLPLAGSLGALNGYELKIQEPPYKPREASVSITKAVKFLKEEHKRNFLFGFA
FMNFAIGVGTVINMVLKEVEFSYLIQISLVNAVGMFIGTLFQPFWGLGDRYGFQYF
LKVCLMIIHAIVILLWTLTPRSFLVYFLOIIGIFVTAGTSQLVFTLYMTAPSSLKT
EAFSVFNSLSNLSLFAGLVASVLVASLENISLPGISAIRLTMFISFPLRASAAYYII
SRMDLGTPOKVDSLQAVKESFSGTVPIRERLNTLNIPRRKR"
complement(3382. .4545)
/locus_tag="TM1061"
complement(3382. .4545)
/locus_tag="TM1061"
/note="similar to percent identity: 0.00; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD36137.1"
/db_xref="GI:4981603"
/translation="MRSMTDVRRYWSILRKICEMPLELCASDRLKSMPEVGKSEE
RRKFTLELGRIFCGISPLERESPTDPEERKIATRLSEFAVSLDVATPNCK
DYMFKEGROPPLVDALFVEAILRAPKYLWEDLDVSTKRLIRELKATRKIIEYPSNW
LLFSAMLETFFEFAGEWDSSTKVDLLKNVESWYKGDGAYGDGPFPRWDYNSFYIYP
MTIDVLRIISEKTEWKELYVKVLRQRAYAVLERMISPEGTPIIGRSITYRTAVF
HLISQLSLHLPLASLPAQVRCALTAVLRIFENPSTFDENGMLKIGVIGSQPSLGE
EYITTGSLYCTVFLPLGLPTSDPFRDPCCKWTKKVMKEGEDVADPRALIED"
complement(4542. .6233)
/locus_tag="TM1062"
complement(4542. .6233)
/locus_tag="TM1062"
/note="similar to PID:642973 PID:14446 PID:412358 percent
identity: 57.73; identified by sequence similarity;
putative"
/codon_start=1

/transl_table=11
/product="beta-glucuronidase"
/protein_id="AAD36143.1"
/db_xref="GI:4981609"
/translation="MVRPQRNKRFFILLNGVWNLEVTSKDRPIAVPGSWNEQYQDLC
YEEGPFYKTTFPYRKELSOXHIRLFPAAVNTDCVFLNGEKGVENHIEYLPFEDVT
GFKVSGENELRVVENRLKVGGFSPKVPDSGTHTVGFSGFPANPDFPYGIIIRPV
LIEFTDARILIDIWDTSESEPEKKIKVKVKEIVSEBAVQEMTIKLEEEKIRTS
NRFVEGEFFILENARFWSLEDPYLYPLKVELEKDEYTLDIGRTISWBEKRLYLNGKPV
FLKFGFGHEEFVLGGGTFFYPLMIKDRNLLKWINANSFRTSHPYSEEWDLADRLGI
LVIDEAPHYGTRTHYNPETQKIAEDNIRMRIDRHNKHPVIMWSVANEPSNHPDAE
GFFKALYETANEMDRRPVVMVSMMDABDERTDVALKFEDIVCVNRYGYWYIYQRI
EEGLQALEKDIIEELYARHRKPIFVTERGADAIAGIHYPDPQMFSEEQAELEVEKTIRL
LKQDYIIGTHWAFADFKTPQNVRRPIINHKGVFTRDRQPKLVAHVLRRLMSEV"
complement(6284. .7309)
/locus_tag="TM1063"
complement(6284. .7309)
/locus_tag="TM1063"
/note="similar to SP:P42065 PID:677944 GB:AL009126 percent
identity: 75.31; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD36138.1"
/db_xref="GI:4981604"
/translation="MALLIEVKNLKXYPPIIKRGFRKRVGYLKAVDGISPHIEEGETL
GLVSGCGCTTTAKLIMRGIEPTGEIVLNMDEKVDITKISEKREKGVRRFLQW
IPQDPYSSLNPRMTVVDIIAEPLVNVGVIKKEEIDAIVSDLLRAVGIRPEYMORYPH
AFSGGQRQRIAIRAIALKPKLVICDEPTSLDVSVAQIINLKQKEYNLTLYFI
SHDLGVEHITNRVAVMYVGRIVEIAETBELFSSPRHPYTEALLSAVPKDPKRKRKK
AQLTGEVPDPSNPSCGYFHRPCPYAKAICKEVYPDLRNVTQDQNPHLVACHPADSLK
LEGVGTW"
complement(7296. .8294)
/locus_tag="TM1064"
complement(7296. .8294)
/locus_tag="TM1064"
/note="similar to SP:P42064 PID:677943 GB:AL009126 percent
identity: 72.95; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, ATP-binding
protein"
/protein_id="AAD36139.1"
/db_xref="GI:4981605"
/translation="MSTLLQIKNLRTYFFTDEGVKAVDGVSEIEBERGRTLGVVGESE
CGKSVTARSIIKILSTAGRIVSGEILVNMDOGMDLVKFSKEIRKVRGRHIAMIFQE
PMAAFSPVYTIQDQTEGMIVHFGITKQEARERAVELIRVGI PKPEKMIIDSYPEYS
GGMKQRAMIAMALSCNPRLLIADERTTALDVTIQAOVLDLKDLQOEYKMAIMMITHN
MGVVAEMADHVVMYIGRVESAPVELEFYNPKHPTSLIIRSIPIVYGRVERLEVIE
GDVDPDRNMPKGCRRFHRPCPYMMKGI CDEREPEVEVGEHRVSCPLYGGEKDGAS"
complement(8306. .9409)
/locus_tag="TM1065"
complement(8306. .9409)
/locus_tag="TM1065"
/note="similar to GB:AE00657 percent identity: 67.80;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oligopeptide ABC transporter, permease protein"
/protein_id="AAD36140.1"
/db_xref="GI:4981606"
/translation="MKKREQMEEKYYASQWQLIMRRFRKHLAVIGGVLLIIYYFA
IFCEFPAPYDPPKNYRYPYAPPRIFHEHEGKFGFVYGYTVLLETIRIYKED
KSKIFRIKPFVRGDEYKFWGIWKTNIHFIGVEDGHMFLGTDLSGRDMLSRITYGARI
STSIGLGVFLSFIGVALIGISGYEGGAVDNFIORTIEIIKSIPPIPLMLALSALP
QMWPLRVLYFVITISLTGMTDLARVVRFLSLREEDFVMAAKFGEASARIIFPRH
MLPSFMSHLIASITSLIPGMTIGERSLFIGLGRPVISWGVLLQEAQNLTVVALYYP
WLLIPVVFVITTVLCFNFGDGLRDAADPYANM"
complement(9420. .10403)
/locus_tag="TM1066"

CDS		complement(9420..10403)
		/locus_tag="TM1066"
		/note="similar to SP:P42062 PID:677946 GB:AL009126 percent identity: 62.93; identified by sequence similarity; putative"
		/codon_start=1
		/transl_table=11
		/product="oligopeptide ABC transporter, permease protein"
		/protein_id="AAD36141.1"
		/db_xref="GI:4981607"
Query Match		99.9%; Score 1688; DB 1; Length 12583;
Best Local Similarity		99.9%; Pred. No. 0;
Matches 1688;		Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	1	ATGTTAAGACCGCGAAAGAAAGAGATTATCTTATCTTGAATGAGTTTGAAT 60
DB	6233	ATGTTAAGACCGCGAAAGAAAGAGATTATCTTATCTTGAATGAGTTTGAAT 6174
QY	61	CTTGAAGTAACCGAAGAACAGACCAATCGCCGTTCTGGAAGCTGGAATGAGCAGTAC 120
DB	6173	CTTGAAGTAACCGAAGAACAGACCAATCGCCGTTCTGGAAGCTGGAATGAGCAGTAC 6114
QY	121	CAGATCTGTGCTACGAAGAGGACCTTCACTCAAAACCACTTCTACGTTCCGAAG 180
DB	6113	CAGATCTGTGCTACGAAGAGGACCTTCACTCAAAACCACTTCTACGTTCCGAAG 6054
QY	181	NAACCTTCAAAAAACATCAGACTTACTTTGCTGCGGTGAACAAGGACTGCGAGGTC 240
DB	6053	GAACCTTCAAAAAACATCAGACTTACTTTGCTGCGGTGAACAAGGACTGCGAGGTC 5994
QY	241	TTCCTCAACGAGAGAAAGTGAGAGAAATCACATTAATACCTTCCCTTGAAGTAGAT 300
DB	5993	TTCCTCAACGAGAGAAAGTGAGAGAAATCACATTAATACCTTCCCTTGAAGTAGAT 5934
QY	301	GTGACGGGGAAGTGAATCCGAGAGAAAGAACTCAGGGTGTGTTGAGAACAGATTG 360
DB	5933	GTGACGGGGAAGTGAATCCGAGAGAAAGAACTCAGGGTGTGTTGAGAACAGATTG 5874
QY	361	AAAGTGGAGAGATTTCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGATTGTTT 420
DB	5873	AAAGTGGAGAGATTTCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGATTGTTT 5814
QY	421	GGAAGTTTTCCACCTGCAAACTTGCATCTTCCCTACGCTGAATCATTAAGCCTGTT 480
DB	5813	GGAAGTTTTCCACCTGCAAACTTGCATCTTCCCTACGCTGAATCATTAAGCCTGTT 5754
QY	481	CTGATAGAGTTCAAGACCAAGCGAGGATCTCGACATCTGGGTGACACAGATGAGTCT 540
DB	5753	CTGATAGAGTTCAAGACCAAGCGAGGATCTCGACATCTGGGTGACACAGATGAGTCT 5694
QY	541	GAAACCGAGAGAACTTGAAAAAGTGAAGTAGAAGTCTCAGAAGAGCGGTG 600
DB	5693	GAAACCGAGAGAACTTGAAAAAGTGAAGTAGAAGTCTCAGAAGAGCGGTG 5634
QY	601	GGAAGAGAGATGAGATCAAACTTGGAGAGAGAGAAAAAGATTAGAATCCAAACAGA 660
DB	5633	GGAAGAGAGATGAGATCAAACTTGGAGAGAGAGAAAAAGATTAGAATCCAAACAGA 5574
QY	661	TTGCTGAAAGGGAGTTTCTCGAAAAACGCCAGTTCTGAGCCTCGAAGTCCATAT 720
DB	5573	TTGCTGAAAGGGAGTTTCTCGAAAAACGCCAGTTCTGAGCCTCGAAGTCCATAT 5514
QY	721	CTTTATCTCTCAAGGTGAAGTGAATAAAGACGAGTACACTCTGAGCATCGAATCAGA 780
DB	5513	CTTTATCTCTCAAGGTGAAGTGAATAAAGACGAGTACACTCTGAGCATCGAATCAGA 5454
QY	781	ACGATCAGCTGGAGCGAGAGAGGCTCTATCTGAACGGGAACTCTTTTGAAGGGC 840
DB	5453	ACGATCAGCTGGAGCGAGAGAGGCTCTATCTGAACGGGAACTCTTTTGAAGGGC 5394
QY	841	TTTGAAGAAGCAGAGAAATTCCTCCGTTCTGGGGCAGGGCACTTTTATCCATTGATGATA 900

DB	5393	TTTGAAGAAGCAGAGAAATTCCTCCGTTCTGGGGCAGGGCACCTTTTATCCATTGATGATA 5334
QY	901	AAAGACTTCAACCTTCTGAAGTGATCAACGGGAATTTCTTCAGGACCTCTGACTATCCT 960
DB	5333	AAAGACTTCAACCTTCTGAAGTGATCAACGGGAATTTCTTCAGGACCTCTGACTATCCT 5274
QY	961	TACAGTGAAGGTGGCTGATCTTTCGCGACAGACTCGGAATCCTTGATAGCGAAGCC 1020
DB	5273	TACAGTGAAGGTGGCTGATCTTTCGCGACAGACTCGGAATCCTTGATAGCGAAGCC 5214
QY	1021	CCGACGTTGGTATCACAAGGTACCCTACTACATCCCGAGACTCAGAAGATAGCAGAAGAC 1080
DB	5213	CCGACGTTGGTATCACAAGGTACCCTACTACATCCCGAGACTCAGAAGATAGCAGAAGAC 5154
QY	1081	AACATAAGAGATGATCGACAGACACAAGAACCATCCCACTGTGATCATGTGAGTGTG 1140
DB	5153	AACATAAGAGATGATCGACAGACACAAGAACCATCCCACTGTGATCATGTGAGTGTG 5094
QY	1141	GCGAACGAACGAGATCCCAACCATCCAGACCGGAGGGTTTCTCAAGCCCTTTATGAG 1200
DB	5093	GCGAACGAACGAGATCCCAACCATCCAGACCGGAGGGTTTCTCAAGCCCTTTATGAG 5034
QY	1201	ACTGCCAATGAATGGATCGAACACAGCCCGCTGTGTCATGTGTGAGCATGATGACGCCACA 1260
DB	5033	ACTGCCAATGAATGGATCGAACACAGCCCGCTGTGTCATGTGTGAGCATGATGACGCCACA 4974
QY	1261	GACGAGAGAACAGAGACGTTGCGGCTGAAGTACTTGACATCTGTGTGAACAAGTAC 1320
DB	4973	GACGAGAGAACAGAGACGTTGCGGCTGAAGTACTTGACATCTGTGTGAACAAGTAC 4914
QY	1321	TACGGCTGTTACATCTATCAGGGAAGGATAGAAAGACCTTCAAGCTCTGGAAGAAAGAC 1380
DB	4913	TACGGCTGTTACATCTATCAGGGAAGGATAGAAAGACCTTCAAGCTCTGGAAGAAAGAC 4854
QY	1381	ATGAAGAGCTCTATGCAAGGCAAGAAAGCCCATCTTTGTCAACAGAAATTCGGTGGGAC 1440
DB	4853	ATGAAGAGCTCTATGCAAGGCAAGAAAGCCCATCTTTGTCAACAGAAATTCGGTGGGAC 4794
QY	1441	GCGATAGCTGGCATCCACTACGATCCACTCAATGTTCTCGAAGAGTACCAAGCAGAG 1500
DB	4793	GCGATAGCTGGCATCCACTACGATCCACTCAATGTTCTCGAAGAGTACCAAGCAGAG 4734
QY	1501	CTCGTTGAAAAAGCAGATCAGGCTCCTTTGAAAAAGACTACATCATCGGAACACACGTG 1560
DB	4733	CTCGTTGAAAAAGCAGATCAGGCTCCTTTGAAAAAGACTACATCATCGGAACACACGTG 4674
QY	1561	TGGGCTTTTGAGATTTTAAGACTCCTCAGAAATGTGAGAAAGCCATTCTCAACCAAG 1620
DB	4673	TGGGCTTTTGAGATTTTAAGACTCCTCAGAAATGTGAGAAAGCCATTCTCAACCAAG 4614
QY	1621	GGTGTTCACAAAGAGACAGACCAACCCAACTCGTGTCTCATGTACTGAGAGAAGCTGTGG 1680
DB	4613	GGTGTTCACAAAGAGACAGACCAACCCAACTCGTGTCTCATGTACTGAGAGAAGCTGTGG 4554
QY	1681	AGTGAGGTT 1689
DB	4553	AGTGAGGTT 4545

RESULT 3
 AF012423 2153 bp mRNA linear MAM 09-SEP-1999
 LOCUS
 DEFINITION Felis catus beta-glucuronidase (gusb) mRNA, complete cds.
 ACCESSION AF012423
 VERSION AF012423.1 GI:4102550
 KEYWORDS
 SOURCE
 ORGANISM
 Felis catus (cat)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
 REFERENCE
 1 (bases 1 to 2153)
 Fyfe,J.C., Kurzhals,R.L., Lassaline,M.E., Henthorn,P.S., Alur,P.R.,
 Wang,P., Wolfe,J.H., Giger,U., Haskins,M.E., Patterson,D.F.,

TITLE	Sun, H., Jain, S. and Yuhki, N.
JOURNAL	Molecular basis of feline beta-glucuronidase deficiency: an animal model of mucopolysaccharidosis VII
PUBMED	Genomics 58 (2), 121-128 (1999)
AUTHORS	Fyfe, J.C., Kurzhals, R.L. and Henthorn, P.S.
REFERENCE	Submitted (03-JUL-1997) Microbiology, Michigan State University, 413 Giltner Hall, East Lansing, MI 48824, USA
JOURNAL	
FEATURES	
source	location/Qualifiers
	1. .2153
	/organism="Felis catus"
	/mol_type="mRNA"
	/db_xref="taxon:9685"
	/tissue type="liver"
	1. .2153
	/gene="GUSB"
	24. 1979
	/gene="GUSB"
	/EC_number="3.2.1.31"
	/note="lysosomal hydrolase; lysosomal exooligosaccharidase in stepwise degradation of glucuronic acid-containing oligosaccharide sidechains of proteoglycans"
	/codon_start=1
	/product="beta-glucuronidase"
	/protein_id="A001498.1"
	/db_xref="GI:4102551"
	/translation="MLRGPAAVWMAALGPLLMACGLALRGMLYPRESPSRERKELNGL WSFRADFSENRRQGFQQWRTPLEBSPTLDMPVSSFENDVGQRQLRSFVGVWG REATLPORWTODIGTRVLRIGSAHYAIWMVNGYHVAHEGHILPFEADISKLVSQ PLASCRITIAINNTLPHTLPGLTIIQTDSKYPRGYFVNQINFDFNYAGLHRPVL LTPTTYIDDTISTSVNQDTGLVDYQIFVEGEHRLBRLDEBKVAQGTGR GQLQVENAHLMWPYLMEHPAYLYSLBVRLLTAOTAGSVSDFTLPLVGITRYAAVTEHQ FLINGKPEFYFGVNKHEDADIRKGFDWPLVKDENLLRWLGANAFRTSHYPVAEVM QLCDRYGIVVIDESPGVIVLVESYSNSVLQHLEWEELVRDKNHPAVVMWSVANE PASFLKPAGYFKTLIAHTKALDPSPVPFTVNSYEADIGA PYDVICVNSYSYWH DYGHEVIOQLATOFENMYRTYQKPIIQSEYGADTIAGFHQDPIMFSEYQGLLE OYHLVLDOKRKEYVVELIMNFADFETNQS PQRVGNKKGI FTNRQPKGAFLRLER YMKLANETRYPMWSAVKSQCLENSPFTL"
ORIGIN	
Query Match	11.5%; Score 194; DB 4; Length 2153;
Best Local Similarity	53.1%; Pred.No. 7,4e-41;
Matches 494; Conservative	0; Mismatches 415; Indels 21; Gaps 3;
Gy	751 GACGAGTACACTCTGGACATCGGAATCAGAAGCATCAGCTGGACGAGAAGGCTCTAT 810
Dp	972 GACTTCTACACTCTCCCCGGTGGAATTCCGACGCTGGCTGTCA CAGACACAGTTCCTC 1031
Gy	811 CTGAACGGGAACCTGTCTTTTGAAGGCTTTGGAAGCACAGAATTCGCCGTTCTG 870
Dp	1032 ATCAATGGGAACCTTTCTATTCCACGGGGTCAACAAGCAGAGATGCAGATATCCGA 1091
Gy	871 GGGCAGGGCACCTTTATCCATTGATGATAAAGACTTCAACTTCTGAAGTGATCAAC 930
Dp	1092 GGGAAAGGCTTTGACTGGCCACTGCTGTGAAGACTTCAATTGCTTCGCTGGCTCGG 1151
Gy	931 GC GAATTTCTTCAGGACCTTCACCTATCTTACAGTAGAAGCTGGCTGAATCTTGCCGAC 990
Dp	1152 GCCAACGCCCTCCGACACCAAGTCATAACCCTACGCCGAGGAGGTGATGCACCTGTGAC 1211
Gy	991 AGACTCGAATCCTTGTGATAGACGAAGCCCGCACGTTGGTATCACAAGGTACCA---C 1047
Dp	1212 CGCTATGGAGATCGTGTCATCGACGAGAGTGCCCGGTGTGGGATCGTGCTGGTCCAAGAC 1271
Gy	1048 TACAATCCCGAGACTCAGAAGATAGCAGAAGCAACATAAAGATGATGCAGACAC 1107
Dp	1272 TACAGCAACGTGTCTTGACAGACCACTGGAAGTGATGAGGAAGCTGGTCCGACGGAC 1331
Gy	1108 AAGAACCATCCAGTGTGATCATGTGAGTGTGGCAACGACCAAGATCCACCATCA 1167

Db	1332	AAGAAATCACC	CGGCTGTTGTGATGTGCTGTGTGGCCAATGAGCCCGCTTCCTCTGAAA	1391	
Qy	1168	GACGCGAGGGTTTCTTCAAAGCCCTTTATGAGACTGCCAATGAATGAATCGAACACGC	1227		
Db	1392	CCTGCCGGTTACTACTTCAAAGACGCTGATTTGCTCACACCAAGCCTTGGATCCCTCCCGG	1451		
Qy	1228	CCCGTTGTCAATGTGAGCATGATGAGCGCACGACGAGAGAAACAAGAGACGTGGCGCTG	1287		
Db	1452	CCCGTGAACCTTTGTGACCACTCCAACTATGAAG-----CAGACCTGGGGCGC	1499		
Qy	1288	AAGTACTTTCGACATCGTCTGTGTGAACAGGTACTACGGCTGTGATCATTCAGGGAAG	1347		
Db	1500	CCGTATGTGACGTGATCTGTGTGAATAGTACTACTCTTGTATCATGACTATGTGAC	1559		
Qy	1348	ATAGAGAAAGGACTTCAAGCTCTGGAATAAAGACATAGAGAGCTCTATGCAAGGACAG	1407		
Db	1560	ATGAGAGTGAATTCAGCTGACGTGGCAACCCAGTTTGAAACTGTATAGGACCTATCAG	1619		
Qy	1408	AAGCCCATCTTTGTCAACAGATTCCGTTGGCGGACGGATAGCTGGCATCCACTACGATCCA	1467		
Db	1620	AAACCAATTAATCCAGAGCGAGTACGAGACAGACACCATTTGACAGGTTTCAACGAGACCA	1679		
Qy	1468	CCTCAATATGTTCTCCGAAAGTACCAAGCAGAGCTCGTTGAAAGACGATCAGGCTCCTT	1527		
Db	1680	CCTCTGATGTTTCAGCGAGGATACCAAGAAAGTCTGTGAGACAGTATCATTTGGTTCTG	1739		
Qy	1528	TTGAAA-----AAGACTACATCATCGGAACACACGTTGGGCTTTGACAGATTTTAAAG	1581		
Db	1740	GATCAAAAACGCAAGAATACGTGTGCGAGAGCTCATCTGGAACCTTGGCGATTTTATG	1799		
Qy	1582	ACTCCTCAGAAATGTGAGAAGACCCATTCTCAACCAAGGGGTGTTTCAACAAGAGACAGA	1641		
Db	1800	ACTAACCACTGACCGCAGAGAGTAAATGGGGAATAAAAAGGGATCTTTCACGCGCAGAGA	1859		
Qy	1642	CAACCCAAACTCGTTGCTCATGTACTGAGA	1671		
Db	1860	CAACCAAAAGGGGACGCTTCTTTTGGCA	1889		
RESULT 4	AF012424	2014 bp	mRNA	linear	MAM 09-SEP-1999
LOCUS	AF012424				
DEFINITION					Felis catus cell-type fibroblast mutant beta-glucuronidase (GUSB)
ACCESSION	AF012424				
VERSION	AF012424.1	GI:4102552			
KEYWORDS					
SOURCE					
ORGANISM					Felis catus (cat)
REFERENCE					Felis catus
AUTHORS					Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
TITLE					1 (bases 1 to 2014)
JOURNAL					Fyfe, J.C., Kurzhals, R.L., Lassaline, M.E., Henthorn, P.S., Alur, P.R., Sun, H., Jain, S. and Yuhki, N.
MEDLINE					Wang, P., Wolfe, J.H., Giger, U., Haskins, M.E., Patterson, D.F., Molecular basis of feline beta-glucuronidase deficiency: an animal model of mucopolysaccharidosis VII
PUBMED					Genomics 58 (2), 121-128 (1999)
AUTHORS					10366443
TITLE					2 (bases 1 to 2014)
JOURNAL					Fyfe, J.C., Kurzhals, R.L. and Henthorn, P.S.
MEDLINE					Direct Submission
PUBMED					Submitted (03-JUL-1997) Microbiology, Michigan State University, 413 Giltner Hall, East Lansing, MI 48824, USA
FEATURES					location/Qualifiers
source					1..2014
					/organism="Felis catus"
					/mol type="mRNA"
					/db xref="taxon:9685"
					/cell type="fibroblast"
					1..2014
					/gene="GUSB"
					gene

CDS		/allele="MPS VII mutant"	2. .1957	/gene="GUSB"	/allele="MPS VII mutant"	/EC_number="3.2.1.31"	/codon_start=1	/product="mutant beta-glucuronidase"	/protein_id="AAD01499.1"	/db_xref="GI:4102553"	/translation="MLRGPAAVMALGLPLLWACGLALRGMLYPRESRRKELNGLWSFRADFSENRRQGFEEQQMYRTPIRESPTLDMPVSSFNDVGODRLRSFYGVWWE REALTPORMTQDLGRVLRIGSAHYAIVMWNVHVAHEHGGLPFEADI SKLVOSG PLASCRITIAINNTLPHTLPPTGILLYQTDTSKYPEKYFVQNINFDFENYAGHRPVL LYTPTTYIDDIITISTSVNQDGTGYDQIFVEGEHHQLFVRLLDEBGXVAVQSTGR GLOVPENAHLMWPYLMEHPAYLSLEVRLLTAQTAAGSVSDEFYTLPEGIRIVATEHQ FLINGKPFYFHGVNKHDADIRKGRDWPLLVKDNLRLWLGNANAFRTSHYPAAEEM QLCDRYGIVIDESPGVGILVLESYSNSVLÖHMLEMBELVRDKNHPRVAMWSVANE PASFLKPAGYYFKTLIAHTKALDPBRPTFTVNSNYEADLGAPYVDVICVNSYYSWH DYGHMEVIOLOLATQEFENMYRQRPILIOSEYGADTIAGFHQDPPLMFSEEQKGLE OYHLVIDOKREYVVGELLIMWFNFADPMTNQSQRVMGNKKGITRQRPKGAFLLRER YWKLNETRYPMSAVKSQCLENSPFTL"
Query Match	11.4%;	Score 192.4;	DB 4;	Length 2014;							
Best Local Similarity	53.0%;	Pred. No. 2e-40;									
Matches 493;	Conservative 0;	Mismatches 416;	Indels 21;	Gaps 3;							
Db	751	GACGAGTACACTCTGGACATCGGAATCAGAAGCATCAGCTGGGACGAGAAGAGGCTCTAT	810								
Db	950	GACTTCTACACTCTCCCCGTGGGATTCCGACGGTGCGTCACAGACGACCAAGTTCCTC	1009								
Qy	811	CTGAACGGGAACCTGTCTTTTGAAGGGCTTTGGAAGCACGAGAATTCCCCTGTCG	870								
Db	1010	ATCATAGGGAACCTTCTATTCCACGGGGTCAACAAGCAAGAAGATGCAGATATCCGA	1069								
Qy	871	GGGACAGGCACTTTTATCCATTGATGATAAAGACTTCAACTTCTGAAGTGATCAAC	930								
Db	1070	GGGAAGGGCTTGACTGGCCACTGCTGTGAAGACTTCAATTGCTTGGCTGCCG	1129								
Qy	931	GCGAATCTTTCAGGACCTCTCACTATCCTTACAGTGAAGAGTGGCTGGAATCTTGCCGAC	990								
Db	1130	GCCAACGCCCTCCGCCACCAGTCACTAACCCCTACGCGGAGGAGTGATGCAGCTGTGAC	1189								
Qy	991	AGACTCGGAATCCTTGTGATGACGAAGCCCCGACGTTGGTATCACAAGTACCA---C	1047								
Db	1190	CGGTATGGGATCGTGGTCACTCGACGAGAGTCCCGGTGTGGGACATCGTGTGCGAGAGC	1249								
Qy	1048	TACAAATCCCGAGACTCAGAAGATAGCAGAAGACAACATTAAGAAGATGATGCACAGACAC	1107								
Db	1250	TACAGCAACGTGTCTGTCAGACCAACCTGGAGGTGATGGAGAAGCTGTGTGGCAGGGAC	1309								
Qy	1108	AAGAACCATCCCACTGTGATCATGTGGAGTGTGGCGAACCAGACCAAGTCCCAACCATCCA	1167								
Db	1310	AAGAATCACCCGGCTGTGTGATGTGTGTGTGTGGCCAATGAGCCCGCTTCCCTCTGAAA	1369								
Qy	1168	GACGCGGAGGGTTTCTTCAAAGCCCTTATGAGACTGCCAATGAATGATGGAACACGC	1227								
Db	1370	CCTGCCGGTACTACTTCAAGACGCTGATTGCTCACACCAAGCCTTGATCCCTCCCGG	1429								
Qy	1228	CCCGTTGTCATGGTGAAGCATGATGAGCAGCACGACGAGAGAACAAAGAGAGTGGCGCTG	1287								
Db	1430	CCCGTGAACCTTGTGACCAACTCCAACATATGAAG-----CAGACCTGGGGGCG	1477								
Qy	1288	AAGTACTTGACATCGTCTGTGTGAACAGGTACTACGGCTGGTACATCTATAGGGAGG	1347								
Db	1478	CCGTATGTGACGCTCATCTGTGTGAATAAGTACTACTCTTGGTATCATGACTATGTGCAC	1537								
Qy	1348	ATAGAAGAAGACTTCAAGCTCTGGAAGAAAAAGACATGAAGAGCTCTATGCAAGGCACAGA	1407								
Db	1538	ATGAGAGGTGATTACAGCTGCAGCTGGCAACCCAGTTTGAGAACTGGTATAGCACTATACAG	1597								
Qy	1408	AAGCCATCTTTGTCAAGAATTGGTGGGACGCGATAGCTGGCATCCACTACGATCCA	1467								

Db	1598	AAACCAATTAATCCAGAGCGAGTACCGGACAGACACCACTTGCAGGGTTTCACCAAGACCCA	1657
Qy	1468	CCTCAATGTTCTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAAGCATCAGGCTCCTT	1527
Db	1658	CCTGTGATGTTTCAGCGAGGAATACCAAGAAAGGCTGCTGAGACAGTATCATTTGGTCTG	1717
Qy	1528	TTGAAA-----AAAGACTACATCATCGGAACACACAGCTGTGGGCGCTTTGCAGATTTAAG	1581
Db	1718	GATCAAAAACGCMAAGAAATACGTGGTTCGAGAGAGCTCATCTGGAACITTTGCCGATTTATG	1777
Qy	1582	ACTCCTCAGAATGTGAGAGACCCCATTTCTAACCCACAGGGGTGTTTTCACAAGAGACAGA	1641
Db	1778	ACTAACCACTCACCAGCAGAGAGTAATGGGGGAATAAAAAGGGGATCTTCACGCCACAGAGA	1837
Qy	1642	CAACCCAACTCGTGTCTCATGTACTGAGA	1671
Db	1838	CAACCAAAAGGGGACAGCGTTCCTTTGGCA	1867
RESULT 5			
LOCUS	AF019759	2199 bp	mRNA linear MAM 30-MAR-1998
DEFINITION	AF019759	Canis familiaris beta-glucuronidase (GUSB) mRNA, complete cds.	
ACCESSION	AF019759		
VERSION	AF019759.1	GI:2425090	
KEYWORDS			
SOURCE		Canis familiaris (dog)	
ORGANISM		Canis familiaris	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.	
AUTHORS		1 (bases 1 to 2199)	
TITLE		Ray, J., Bouvet, A., Desanto, C., Fyfe, J. C., Xu, D., Wolfe, J. H., Aguirre, G. D., Patterson, D. F., Haskins, M. E. and Henthorn, P. S. Cloning of the canine beta-glucuronidase cDNA, mutation identification in canine MPS VII, and retroviral vector-mediated correction of MPS VII cells	
JOURNAL		Genomics 48 (2), 248-253 (1998)	
MEDLINE		98190525	
PUBMED		9521879	
REFERENCE		2 (bases 1 to 2199)	
AUTHORS		Henthorn, P. S., Fyfe, J. C. and Bouvet, A. B.	
TITLE		Direct Submission	
JOURNAL		Submitted (15-AUG-1997) Section of Medical Genetics, University of Pennsylvania School of Veterinary Medicine, 3900 Delancey St., Philadelphia, PA 19104, USA	
FEATURES		Location/Qualifiers	
source		1. .2199	
gene		/organism="Canis familiaris"	
CDS		/mol_type="mRNA"	
		/db_xref="taxon:9615"	
		1. .2199	
		/gene="GUSB"	
		63. .2018	
		/gene="GUSB"	
		/EC number="3.2.1.31"	
		/note="lysosomal enzyme"	
		/codon start=1	
		/product="beta-glucuronidase"	
		/protein_id="AAC48809.1"	
		/db_xref="GI:2425091"	
		/translation="MSRGPAGAWVALGPLLMTGGLALEGMLYPRESPSRERKDLGLWSFRADFSDDRQGFEGQWYRAPLRESGPTLDMVPVSSFNDYGQDRQLRSFVGWVWYERATLPRRWSQDPGTRVLRIGSAHYAIWNVGVHVAHEGHLRFEADISKLVSQPLSSCRITLAINNTLTPHTLPBGITVYKTDASKYPKGYFQNTYFDFFPNYAGLHRPVL	
		LYTPTTYIDIDITVTTGVDODGLVDYQIFVQSEHFQLEYYLLDEEGKVAAGTGSQGRLOVPNVHLMWPLYLMHEHPAYLXLSLEVRLLAOMAAGPVSDEYTLPGVIRYAVTERQFLNGKPYFYGWKNKHEDADIRGKGFDMPLVKDFNLRLWGANAFTSHYPYAEEMQLCDRXYIVVIDESPGVGIMLVQYSNVSLQHLEVMGELVRDKNHPVVMWSVANEPTSLKPAAYVYFKTLIAHTKALDPSRPVTFVNSNYEADLGAAPYVDVIVNSYYSWYHDYGMHEVIQQLATFENWRYTKQPIIQSEYGAETIAGFHDDPPLMFSEBYQKGLLEQYHLVLDQRKEYVVGELIIMWADFMTDQSPQRAVGNRKGIETFRQRPKAAAFLLRERYKLANETGHHRSAAKSQCLENSPFAL"	
ORIGIN			

Query Match		11.3%;	Score 190.2;	DB 4;	Length 2199;
Best Local Similarity		52.8%;	Pred. No. 7.9e-40;		
Matches 492;		Conservative 0;	Mismatches 418;	Indels 21;	Gaps 3;
QY	750	AGACGAGTACACTCTGGACATCGGAATCAGAACGATCAGCTGGACGAGAAAGAGCTCTA	809		
DB	1010	AGACTTCTACTCTCCCGTGGGATTCGACCGTGCCCTCACAGAGCCGACGTTCT	1069		
QY	810	TCTGAACGGGAAACCTGTCTTTTGAAGGGCTTTGAAGCAGAGAAATCCCGTTCT	869		
DB	1070	CATCAACGGGAAACCTTTCTATTCCATGGGGTCAACAAACATGAGATGCCGATATCCG	1129		
QY	870	GGGCAAGGACACCTTTTATCCATTGATGATAAAGACTTCAACCTTGAAGTGATCAA	929		
DB	1130	AGGGAAGGGCTTTGACTGGCCGCTGTGGTGAAGACTTCAACCTGTGGCTGGGG	1189		
QY	930	CGCGAATCTTTACAGACCTCTCACTATCCTTACAGTGAAGAGTGGCTGATCTTGCCGA	989		
DB	1190	CGCGAATGCTTCCGACACGACCTACCCCTACGCGAGAGAGTGATGCACTCTGCGA	1249		
QY	990	CAGACTCGGAATCCTGTGATAGACGAGCCCGCAGCTTGATCAAGTACCA---	1046		
DB	1250	CCGTAATGGGATCGTGTGCATGACGAGAGCCCTGTGTGGGATCATGCTGTCCAGAG	1309		
QY	1047	CTACAATCCCGAGACTCAGAGATAGACGAAGACACACATAAGAATGATCGACAGACA	1106		
DB	1310	CTACAGCAATGTGTCCCTGCAGACCACTGTGAGGTGATGGGGAGCTGTGCTGGGA	1369		
QY	1107	CAAGAACCATCCCACTGTGATCATGTGAGAGTGTGGCGAAGAACCAAGTCCCAACCATCC	1166		
DB	1370	TAAGAAATCAACCATCTGTAGTATGATGTGTGTAGCCCAATGAGCCCACTTCTCTGAA	1429		
QY	1167	AGACGGGAGGGTTCTTCAAGCCCTTATGAGACTGCCAATGAATGATCGAACACG	1226		
DB	1430	GCCTGCTGCTTACTACTCAAGACGCTGATTGCTCACCAAGGCCCTTGACCCCTCCG	1489		
QY	1227	CCCCCTGTCTATGCTGACATGATGACGCGACCAAGAGAGAAAGAGAGAGCTGCGCT	1286		
DB	1490	GCCCGTGAACCTTGTGACCAATTCCAACTATGAAG-----CAGACCTGGGGGC	1537		
QY	1287	GAACTACTTGCACATCGTCTGTGTGAACAGGTACTACGGCTGTGATCTATCAGGGAAG	1346		
DB	1538	GCCGTATGTGACCGTATCTGTGTCAACAGTTACTTCTGTATCAAGACTATGGGCA	1597		
QY	1347	GATAGAAGAAGACTTCAAGCTCTGGAAGAAAGACATAGAAGAGCTTATGCAAGGCAAG	1406		
DB	1598	CATGAGGTGATTACGCTGACGCTGGCCACCGAGTTGAGAACTGTTATAGGACCTAACCA	1657		
QY	1407	AAAGCCCATCTTTGTACAGAAATTCGGTGGGACGCGATAGCTGGCATCCACTACGATCC	1466		
DB	1658	GAAACCAATATCCAGAGCGAGTACGGGGCAGAGACAATTGCAAGGCTTCACACGAGATCC	1717		
QY	1467	ACCTCAATGTTCCTCGAAGAGTACCAAGCAGAGCTGTTGAAAAGACGATCAGGCTCCT	1526		
DB	1718	ACCTCTGATGTTCACTGAGAGTACAGAAAGTCTGCTCGAGCAATACCTTGTGCT	1777		
QY	1527	TTTGAA-----AAAAGACTACATATCGGAACACACGTGTGGGCTTTGCAGATTTAA	1580		
DB	1778	GGATCAGAAACGCAAGAAATATGTGTGGAAGAGCTCATCTGAAATTTGTGATTTAT	1837		
QY	1581	GACTCTCAGAAATGTGAGAAGACCAATCTCAACCAAGGGTGTGTTTCAAGAGACAG	1640		
DB	1838	GACTGACCAAGTACCAAGAGAGCAGTGGGAACAGAAAGGCACTTCACTCGCCAGAG	1897		
QY	1641	ACAACCCAAACTCGTTGCTCATGTAAGTACA	1671		
DB	1898	ACAACCCAAAGCGGCGCTTCTTTTGGCA	1928		

RESULT 6
AF084552
LOCUS AF084552 2155 bp mRNA linear PRI 07-SEP-1998

DEFINITION	Chlorocebus aethiops beta-glucuronidase mRNA, partial cds.
ACCESSION	AF084552
VERSION	AF084552.1 GI:3549608
KEYWORDS	
SOURCE	Cercopithecus aethiops (African green monkey)
ORGANISM	Cercopithecus aethiops
REFERENCE	1 (bases 1 to 2155)
AUTHORS	Vervoot R.
TITLE	Partial cDNA sequence of Cercopithecus aethiops (COS7 cell)
JOURNAL	beta-glucuronidase
REFERENCE	2 (bases 1 to 2155)
AUTHORS	Unpublished
TITLE	Unpublished
JOURNAL	Direct Submission
FEATURES	Submitted (16-AUG-1998) Cell Genetics, Medical Research Council Human Genetics Unit, Crewe Road, Edinburgh EH4 2XU, U.K.
source	Location/Qualifiers
	1..2155
	/organism="Cercopithecus aethiops"
	/mol_type="mRNA"
	/db_xref="taxon:9534"
	/cell_line="COS7"
	<1..1948
	/EC_number="3.2.1.31"
	/function="hydrolase"
	/note="similar to Homo sapiens beta-glucuronidase; beta-D-glucuronoside; glucuronosohydrolase; glucuronidase; glucuronohydrolase"
	/codon_start=2
	/product="beta-glucuronidase"
	/protein_id="AAC34593.1"
	/db_xref="GI:3549609"
	/translation="GLAMAAVILGLIMGCALALDGGMLYPRESQSRERKELDLMSF RADFSNRRRGFEBOYRRLRESGPTLDMPPSSFNDISQWRRLRFVGMVYEREL ILPERMOTDLSTRVYLRLIGSAHAYAVVWNGVHTLEHGGYLPFEADISNLVQGPLS SHVRLTAINNTLSTLPGTIQYLDISKYPKGYLQNTYDFENVAGLQSRVLLY TTPRAYIDITVTGTEHDTGLVNTQISVKSNLFELEVRLLDAENGLVANGTIOGQ LKVPGARLWMPYLMHERPAYLYSLVRLTAQSLGPSDFYTLPGVIRTVAVTESQFL INKRPYFHVGNKHEDADIRGKFPDPLVKDFNLRLWLGANAFRTSHYPYAEVLOM CDRYGVIVIDECPGVGLALPQFPNNVSLQNHMRVMEVVRDKNHPAVVMWSVANEPA SHLESAGYLLKMTVHTKALDPSRPVFTVNSNYADKGAAYVDVCLNSYSWYHDY GHLELIQRLTTOFENWYKTYQKPIIOSEYGAETIYGFHQPDPIMFTFEYQKSLBEQY HVLIDQRRKYVVEGLIWNFADFMTEQSPTRVLGNKKGVFTRRQPKSAFLRLRYW KIANETRYPHSIAKSCLENSPFT"
	2103..2108
	2126..2131
polyA_signal	
polya_signal	
ORIGIN	
Query Match	10.3%; Score 174.4; DB 9; Length 2155;
Best Local Similarity	51.6%; Pred. No. 1.5e-35;
Matches 487; Conservative	0; Mismatches 436; Indels 21; Gaps 3;
QY	737 TGGAACCTGAAAAGACGAGTACACTCTGACATCGGAATCAGACGATCAGCTGGAGC 796
DB	930 TGGGGCTGTGTGACTTCTACACCTCCCTGTGGGATCCGACGTGCTGTCAACG 989
QY	797 AGAAGAGGCTTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGGAAGCAGAG 856
DB	990 AAAGCAGTTCCTCATCAATGGGAAACCTTTCTATTCCACGGTGTCAACAACATGAG 1049
QY	857 AATTCCTCCGTTCTGGGAGGGAACCTTTATTCATTGATGATAAAGACTTCAACCTTC 916
DB	1050 ATGCGACATCCGAGGAGGCTTCGACTGGCCGCTGTGTGAAGACTTCAACCTGC 1109
QY	917 TGAAGTGAATCAACGGAATCTTTCAGAGACCTCTCACTATCCTTACAGTGAAGAGTGC 976
DB	1110 TTCGCTGCTGTGGTCCAAATGCTTCGACACGACCACTACCTTACGCCGAAAGTGC 1169
QY	977 TGATCTTGCAGACGACTCGGAATCCTGTGATAGACGAGCCCGCAGCTTGATACA 1036

Db	1170	TGCAGATGTGTGACCCCTATGGGATTGTGTCATCGATGAGTGTCTCGGCTGGGCTGG	1229
QY	1037	CAAGGTACCA---CTACATCCCGAGACTCAGAAGATAGCAGAACAACATPAGAA	1093
Db	1230	CACGTCCCGCAGTCTTCAACAACGCTGTCCCTGCAGAACACATGGGGGTGATGGAGAA	1289
QY	1094	TGATCGACAGACACAAGAACCATCCAGTGTGATCATGTGGAGTGGCGAACCAG	1153
Db	1290	TGGTGGCGCAGGAGACAAGAACCCCGCGCTCGTGATGTGTGTGGCCAACGACCTG	1349
QY	1154	AGTCCAACCATCCAGACCGCGAGGGTCTTCAAAAGCCCTTTATGACACTGCCAATGAAA	1213
Db	1350	CGTCCCACTAGATCTGCCGCTACTACTTGAAGATGGTGATCACTCACACCAAGCCCT	1409
QY	1214	TGGATCGAACACGCCCCGTTGTCTATGTGTGAGCATGATGACCGCACAGACGAGAACAA	1273
Db	1410	TGGACCCCTCCCGGCTGTGACCTTTGTGACCAACTCCAACCTATGCGAG-----	1457
QY	1274	GAGACGTGGCGCTGAAGTACTTCGACATCGTCTGTGTGAACAGGTACTACGGCTGGTACA	1333
Db	1458	CAGACAAGGGGGCTCCGTATGTGAGCGTGATCTGTTTGAACAGCTACTACTCTTGGTATC	1517
QY	1334	TCTATCAGGGAAGATAGAGAAGACTTCAAGCTCTGAAAAAACAATAGAGAGCTCT	1393
Db	1518	ACGACTATGGGCACCTGGAGTTGATTCAGCGGCAGCTTACCACCCAGTTTGAGAACTGGT	1577
QY	1394	ATGCAAGGCACAGAAAAGCCCATCTTTGTACAGAAATTCGGTGGAGCGGATAGCTGGCA	1453
Db	1578	ATAAGACGTATCAGAAAGCCCATTTATTCAGAGCGAGTATGAGCGGAAACGATTGTTGGGT	1637
QY	1454	TCCACTACGATCCACTCAAAATGTTCTCCGAAGAGTACCAGCAGAGCTCGTTGAAAAA	1513
Db	1638	TTCAACGAGACCCGCTCTGATGTTTCACTGAAGAGTACCAGAAGAGTCTGCTAGAGCAGT	1697
QY	1514	CGATCAGGCTCCTTTTGA AAAA-----AGACTACATCATCGGAACACACAGTGGGCT	1567
Db	1698	ACCATGTGTTCTTGATCAAAAAACGAGAAAGTACGTGTTGGAGAGCTCATCTGGAATT	1757
QY	1568	TTGCAGATTTTAAAGACTCCTCAGAAATGTGAGAAAGCCCATTTCAACCAAGGGTGT	1627
Db	1758	TTGCCGATTTTCATGACTGAACAGTCAACCGACGAGAGTGTGGGGAATAAAAAGGGGTCT	1817
QY	1628	TCACAAGAGACAGACAAACCAACTCGTGTCTCATGTACTGAGA	1671
Db	1818	TCACTCGGCAGAGAACCAAAAAAGTGACGCTTCTTTGGCA	1861

RESULT 7
CQ575703 2128 bp DNA linear PAT 02-FEB-2004
LOCUS CQ575703 Sequence 3461 from Patent WO0171042.
DEFINITION CQ575703
ACCESSION CQ575703
VERSION CQ575703.1 GI:41639115
KEYWORDS
SOURCE Drosophila sp.
ORGANISM Drosophila sp.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1
AUTHORS Venter,J.C., Adams,M., Li,P.W. and Myers,E.W.
TITLE Detection kits, such as nucleic acid arrays, for detecting the
expression of 10,000 or more Drosophila genes and uses thereof
JOURNAL Patent: WO 0171042-A 3461 27-SEP-2001;
PE Corporation (NY) (US)
FEATURES
Location/Qualifiers
1..2128
/organism="Drosophila sp."
/mol_type="unassigned DNA"
/db_xref="taxon:7242"

ORIGIN
Query Match 10.3%; Score 174; DB 6; Length 2128;

Beet Local Similarity 51.6%; Pred. No. 1.9e-35;				
Matches 492; Conservative 0; Mismatches 435; Indels 27; Gaps 3;				
QY	734	AGGTGAACCTGAAAAAGCAGTACACTGTGACATCGGAATCAGAACATCAGCTGG	793	
Db	1104	AGTTGAATCTCTGCAGATTAACCTACCGCTCCAGTGGGCATACGTAAAGTTGG	1163	
QY	794	ACGAGAAGGCTCTATCTGAACGGGAACTGTCTTTTGAAGGCTTTGAAAACAG	853	
Db	1164	ACAACGACGCTGTCTGTGAATGAAAAACCTCTATCTGCGGGATTGACGGCAG	1223	
QY	854	AGGAATTCCTCTCTGGGCAAGGACCTTTATCCATTGATGATMAAGACTTCAACC	913	
Db	1224	AGGACTCCGATATCCGGGAAAGGATGGAATATGCGCTTCTGTAGATTTTAAAC	1283	
QY	914	TTCTGAAGTGATCAACGCGAATCTTTCAGACCTCTCACTATCCTTACAGTAAGAGT	973	
Db	1284	TGCTGAAGTGACTGAGCCAAATGCATATGCACTCTCACTATCCTTATCCGAAGAGT	1343	
QY	974	GGCTGATCTTGCCGACAGACTCGGAATCCTTGATAGACGAAGCCCGCAGCTTGTA	1033	
Db	1344	CAATGACGTTTGGCCGATCAGCATGTATTAATGATTATGACGAATGCC-----TGCTG	1397	
QY	1034	TCACAAGGTACCACTACAATCCCGAGACTGAGAAGATAGCAAGAACAACTAAGAAGAA	1093	
Db	1398	TCAATATAGATATCTTCGAGCCGAGCTACTGAGAAATACATATGCTCGCTGAGCAAC	1457	
QY	1094	TGATGACAGACACAAGAACCATCCAGTGTATCATGTGAGTGTGGCGAACGAACAG	1153	
Db	1458	TGATCCACGGGACAGAAACCATCCAAAGTGTGTGATGTCGTAAGCAGACGCGCA	1517	
QY	1154	AGTCCAACCATCCAGACGGGAGGTTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAAA	1213	
Db	1518	GATCGAACAGCAGGAGGCCCTTAATACTTTGAATTCCTGTGTAACATACTATGAAGAAA	1577	
QY	1214	TGGATGAACAGCGCCCGTGTCTATGTGATGATGATGAGCGCACAGAGAGAGAACAA	1273	
Db	1578	TAGCTCAGCGACGACCTCTAACCGCGCTAATAATGCAACTC-----TTCCA	1625	
QY	1274	GAGACGTGGCGCTGAAGTACTTCGACATCGTGTGTGAACAGTACTACGCTGTGACA	1333	
Db	1626	GTTGCCATTTGGCGCAGTTCTGACATCGTGGGTTTCAATCGCTATACTCTTGTATC	1685	
QY	1334	TCTATCAGGGAAGATAGAGAAGACTTCAAGCTCTGAAAAAGACATAGAGAGCTCT	1393	
Db	1686	AGAACTCGGGCGCTACTGACATGATACTTAACCTGTTGACGATAGAGGCCAGAGTTGGC	1745	
QY	1394	ATGCAAGGACAGAAAGCCCATCTTTGTCAAGAAATTCGGTGGAGCGGATAGCTGCA	1453	
Db	1746	GGGATAGGTTCCGAAAGCCTGTATCCAAATTTAGTACGGGGCGACACTATGAGAGGCA	1805	
QY	1454	TCCACTACGATCCACCTCAATGTCTCCGAAGAGTACCAAGCAGAGCTGT-----	1505	
Db	1806	TGCACCTCACTCCCGCCTTATTTGTGCGAGGAATACCAAGTTGAGCTTCTCTCGCC	1865	
QY	1506	-TGAAGAAGCATCAGGCTCTTTTGA AAAAAGACTACATCATCGGAACACAGCTGTGG	1564	
Db	1866	ATTTCAAGGCTTTTCAGCAGATTACGCGGAAGAGATGTTTATTTGAGAGTTGTTTGA	1925	
QY	1565	CCTTTCAGATTTTAAAGACTCTCAGAATGTGAGAAGACCCATTTCTCAACCAAGG	1624	
Db	1926	ACTTCGCCGATTTTTCGACCGCGCAGACTTATTAACCGCGTGGCGGCAACAAAAGGAG	1985	
QY	1625	TTTTTCAACAAGACAGACAAACCAACTCGTGTCTCATGTACTGAGAAGACTGT	1678	
Db	1986	TCTTTACAAGGAACCGAACCAACCAAGAGTGCTCAACATTTCTTAAAGCGGCGGT	2039	

RESULT 8
BC014142 2182 bp mRNA linear PRI 29-JUN-2004
LOCUS BC014142
DEFINITION Homo sapiens glucuronidase, beta, mRNA (cDNA clone MGC:20659
IMAGE:4662011), complete cds.

ACCESSION BC014142
VERSION BC014142.2 GI:40225978
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2182)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scaetiz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalhus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2182)
Strausberg,R.
Direct Submission
Submitted (10-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
On Dec 19, 2003 this sequence version replaced gi:15559559.
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaapi,R., Maduro,Q.L., Mastello,C., Maderi,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stantiriop,S., Thomas,P.J., Touchman,J.W., Tsurgion,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRAL Plate: 30 Row: C Column: 15
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504222.
location/Qualifiers
1..2182
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:20659 IMAGE:4662011"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NIH_MGC_15"
/lab_host="DH10B-R"
/note="Vector: pOT87"

gene 1..2182
/gene="GUSB"
/note="synonym: MPS7"
/db_xref="locusID:2990"
/db_xref="MIM:253220"
2..1957
/gene="GUSB"
/codon_start=1
/product="glucuronidase, beta"
/protein_id="AAH14142.1"
/db_xref="GI:15559560"
/db_xref="LOCUSID:2990"
/db_xref="MIM:253220"
/translation="MARGSAVAMALGPILLMGALGLGGMLYPQESPSRECKELDGLWSFRADFSNRRRGFEQWYRPLWESGPTVDMPPVPSFNDISODWRLHFVGMWYERVLPERMTQDLRTRVLRIGSAHSYAIVWNGVDTLEHGGYLPFEADISNLVQVGLPSRLRTIAINNTLPPTLP.GGTIOYLTDSKYPKGYFQNTYFDFNYAGLORSVLLYTPPTTYIDITVTTSVEQDSGLVNYQISVKSNI.FKLEVRLLDAENKVVANGTGOFLINGKPFYFHGVNKHEDADIRGKGFDP.LVKDFNLRLWLGANAFRTSHYPAAEVMMCDRGIVVIDECPGVGLALPQFNNVSLHHMNVBEVVRDRDKNPAVMMSVANEPASHLSAGYLLKMYIAHTKSLDPSRPVTVSNSNTAADKAPYDVICLNSYSWYHDYGLHELIQLATQFENMYKKYQKPIIOSEYGAETIAGFHODPLMFTEBYQKSLBQYHLGLDQKRKYVGGELIWNFADFMTEQSPTRVLGNKKGIPTRRQPKSAAPFLRERYWKIANETRYPHSVAKSQCLENSLFT"

ORIGIN
Query Match 10.3%; Score 174; DB 9; Length 2182;
Best Local Similarity 52.2%; Pred. No. 1.9e-35;
Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;
QY 737 TGGAACTTGA AAAAGACGAGTACCTCTGGACATCGGAATCAGAACGATCAGCTGGGACG 796
DB 939 TGGGGCTGTGTGACTTCTTACACACTCCCTGTGGGAGATCCGACTGTGCTGTACCA 998
QY 797 AGAAGAGGCTTATCTGAACGGGAACTGTCTTTTGAAGGCTTTGGAAGCAGAG 856
DB 999 AGAGCCAGTTCATCATATGGGAAACCTTTCTATTCCACGGTGTCAACAAGATGAGG 1058
QY 857 AATTCCCGTTCTGGGGCAGGGACCTTTTATCCATTGATGATAAAGACTTCAACTTC 916
DB 1059 ATGCGGACATCCGAGGGAAGGGCTTGCAGTGGCGCTGTGTAAGACTTCAACTTC 1118
QY 917 TGAAGTGAATCAACGGAATCTTTCAGAGACCTCTCACTATCTTACAGTGAAGTGGC 976
DB 1119 TTGCTGGCTGTGTGCCAACGCTTTCGTACCAAGCCACTACCCCTATGCAGAGAACTGA 1178
QY 977 TGGATCTTCCGACAGACTCGGAATCCTTGTGATAGACGAAGCCCGCAGCTGGTATCA 1036
DB 1179 TGCAGATGTGTGACCGCTATATGGGATGTGTGATCATGATGATGTCCCGCTGGCCTGG 1238
QY 1037 CAAGGTACCA---CTACAATCCGAGACTCAGAGATAGCAGAGAACAACATAGAAGAA 1093
DB 1239 CGTGCAGAGTTCTTCAACAAGTTTCTCTGCATCACCACATGCAAGTGATGAAGAAAG 1298
QY 1094 TGAATCGACAGACAAGAACCATCCAGTGTGATCATGTGAGTGTGGAGGAACGACAG 1153
DB 1299 TGGTGCATAGGACACAAGAACACCCCGCGTGTGATGTGTGTGGCAACGACCTG 1358
QY 1154 AGTCCAACCATCAGACGCGGAGGTTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAAA 1213
DB 1359 CGTCCACACTAGATCTGTGGCTACTACTTGAAGATGTGATCGCTCACACCAATCCT 1418
QY 1214 TGGATCGAACACAGCCCGCTTGTGATGTGAGCATGATGACGCAACGACGAGAACAA 1273
DB 1419 TGGACCCCTCCGCGCTGTGACCTTTGTGAGCACTTAACATATGACG----- 1466
QY 1274 GAGACGTGCGCTGAAGTACTTGCATCGTCTGTGTGAACAAGTACTACGGCTGTACA 1333
DB 1467 CAGACAAGGGGCTCCGTATGTGATGTGATCTGTTGAACAGCTACTACTCTTGGTATC 1526
QY 1334 TCTATCAGGAGGATAGAAGAAGACTTCAAGCTCTGAAAAAAGACATAGAAGAGCTCT 1393

Db 1527 ACGACTACGGGCACCTGGAGTTGATTCAGCTGCAGCTGGCCACCAGTTGAGAACTGGT 1586
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTTGTACAGAAATTCGGTCCGACCGCATAGCTGGCA 1453
Db 1587 ATAGAGAGTATCAGAAAGCCCATTTATTCAGAGCGAGTATGGAGCAGAAACGATTTGCCAGGGT 1646
QY 1454 TCCACTACGATCCACCTCAATGTCTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAAG- 1512
Db 1647 TTCACCAAGATCCACCTCTGATGTTCACTGAAGAGTACCAAGAAAAGTCTGCTAGACAGT 1706
QY 1513 ACGATCAGGCTCCTTTGAAAAA----AGACTACATCATCGGAACACACAGTGGGGCCT 1567
Db 1707 ACCATCTGGGCTCTGGATCAAAAACGAGAAAATACGTGTTGGAGAGCTCATTTGGAATT 1766
QY 1568 TTGCAGATTTTAAGACTCTCAGATGTGAGAAAGCCATTCTCAACCACAAGGGTGT 1627
Db 1767 TTGCCGATTTTCATGATGACACAGTCAACCGACGAGAGTCTGGGGAAATAAAAGGGGATCT 1826
QY 1628 TCACAAGAGACAGACAACCCAACTGCTGCTCATGTACTGAGA 1671
Db 1827 TCACTCGGACAGACACCAAAAAGTGCAAGCGTCTTTGCCA 1870

RESULT 9
BT010073 2321 bp mRNA linear INV 12-AUG-2003
LOCUS Drosophila melanogaster LDI0588 full insert cDNA.
DEFINITION BT010073
ACCESSION BT010073
VERSION BT010073.1 GI:33589549
KEYWORDS FLI cDNA.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster

REFERENCE
AUTHORS Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
Champe,M., Chavez,C., Dorsett,V., Dreesnek,D., Farfan,D., Frise,E.,
George,R., Gonzalez,M., Guarin,H., Kronmiller,B., Li,P., Liao,G.,
Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
Patel,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.
and Celniker,S.
1 (bases 1 to 2321)
Ephydroidea; Drosophilidae; Drosophila.

TITLE Direct Submission
JOURNAL Submitted (12-AUG-2003) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
COMMENT Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our Web site
(http://fruitfly.berkeley.edu) or send email to
cdna@fruitfly.berkeley.edu.

FEATURES
source
1..2321
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
1..2321
gene
/gene="CG2135"

CDS
/db_xref="FLYBASE:FBgn0039874"
64..2124
/gene="CG2135"
/note="Longest ORF"
/codon_start=1
/product="LD10588p"
/protein_id="AA022542.1"
/db_xref="GI:33589550"
/db_xref="FLYBASE:FBgn0039874"
/translation="MKSHMPVCLILGSLQVLAATTENDFLVNAARKELTDSPLFEY
DNVRKPTPKGLLPRDSETRVRLDGMQLVRSDPYDPLQIREKFMALRKTR
EISMPVPASVNDITTDNLRDHYGTWYERTFFVRSWKTMRWLRFSSVHSAVW
INGRNATSHSIGLHPFESEISGLISFGGDNRIYVCDNRLSNRTIPQGSVYKVATDDG
TVPISYTFDFFNVAAGIHRSVHLVTTPLHISDEVTTLQITKEGLRIDYRVWLDAK
EGLQIQPIQLRVQLRDXDGHVAAQOINKAVHGTLLVPMATPWWPYLMSDPGLYNL
OFELFVASNEKELESLODTYRLPVGIRSLSWNDSLLNGKPLVLRGFRHEDSDIRG
KGLDNALLARDNFLKWTGANAYRTSHYPYSESMQFADQHGIMIIDECPAVNIDIFE
POLLENHMSLEOLIHDRDNHPSVAVSVANBPSNKGALKYFEFLVNVYREIAHGR
PLTAIINANSSCHLAFDLIVGFNRVNSWYQNSGRTDMLNLTLIEAQSWDRDFGK
VIOPEYGDTEGMSLPAFIVSEYQVELFSRHFKAFDELRGWFIQGFVWNRADF
RTAQITTRVGGNKKGVFTRNRQPEVAHILRRYFALAKELDMFGLPEDLTVYISKHI
HNEL"

ORIGIN

Query Match 10.3%; Score 174; DB 3; Length 2321;
Best Local Similarity 51.6%; Pred. No. 1.9e-35;
Matches 492; Conservative 0; Mismatches 435; Indels 27; Gaps 3;

QY 734 AGGTGAACCTGAAAAAGACGAGTACACTCTGGACATCGGAATCAGAACGATCAGCTGG 793
Db 1100 AGTTGAATCTCTGCAGAGTACCTACCGCCTCCAGTGGGCATACGTAGCTTAAGTTGG 1159
QY 794 ACGAAGAGGCTCTATCTGAAACGGGAACCTGCTTTTGAAGGGCTTTGAAAAGCAGC 853
Db 1160 ACAACGACAGCTGCTGCTGAATGAAAACCCCTCATCTGCGGGATTGGACGGCAGC 1219
QY 854 AGGAATTCCTCGTCTGGGACGGGCACCTTTATCCATGTATGATAAAGACTTCAACC 913
Db 1220 AGGACTCCGATATCCGCGGAAAGGATTGATATGCGCTTCTTGCTAGAGATTTTAACC 1279
QY 914 TTCTGAAGTGAATCAACGCAATTTCTTCAGAGCCTCTCACTATCCTTACAGTGAAGAGT 973
Db 1280 TGCTGAAGTGAAGTGAAGCAATGATATGCAACCTCTCACTATCCTTATTCGGAAGAGT 1339
QY 974 GGCTGATCTTCCGACAGACTCGGAATCCTGTGTATAGACGAAGCCCCGACGTTGGTA 1033
Db 1340 CAATGCAATTTGCGCATCAGCATGTGATTAATGATTATGACGAATGCC-----TGCTG 1393
QY 1034 TCACAAGGTACCACTACAAATCCCGAGACTCGAAGATAGCAGAAAGACAACATTAAGAA 1093
Db 1394 TCAATATAGTATCTTGCAGCCGACGCTACTGAGAAATCATATGTCCTCGCTGAGCMA 1453
QY 1094 TGATGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGCAACAG 1153
Db 1454 TGATCCACCGGACAGAGAACCAATCCAAAGTGTGTGATGTCGGTAGCCACGAGCCGA 1513
QY 1154 AGTCAACCATCCAGACGCGGAGGGTTTCTTCAAGCCCTTATGAGACTGCCAATGAA 1213
Db 1514 GATCGAACAGCAGAGAGCCCTTAATATCTTGAATCTCTGTAACACTATGTAAGAGAA 1573
QY 1214 TGATCGAACACGCCCGTGTCTATGTGTGATGATGAGCAGCACGAGAGAGAACAA 1273
Db 1574 TAGCTCAGGACGACCTCTAACCGCGGTATAATGCCAACTC-----TTCCA 1621
QY 1274 GAGACGTGGCGCTGAAGTACTTGACATCGTGTGTGAACAGTACTACGCGTGTGACA 1333
Db 1622 GTTGCAATTTGGCGCAGTTTCTGACATCGTGGGTTTCAATCGCTATACTCTTGTATC 1681
QY 1334 TCTATCAGGGAAGATAGAGAAGACTTCAAGCTCTGAAAAAAGACATAGAGAAGCTCT 1393
Db 1682 AGAAGCTGGGCGTACTGACATGATACTTAACCTGTTGACAGATAGAGCGCCAGAGTTGC 1741
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTTGTGACAGAATTCCGTGCGGACGCGATAGCTGCA 1453

Db 1742 GGGATAGGTTTCGAAAGCCTGTATCAATTGAGTACGGCGGACACATATGAGGGCA 1801
Qy 1454 TCCACTACGATCCACCTCAAAATGTTCTCCGAAGAGTACCAAGCAGAGCTCGT----- 1505
Db 1802 TGCACCTCACTTCCCGCTTTATTTGTCGGAAGATACCAAGTTGAGCTCTTCTCGCC 1861
Qy 1506 -TGAAAAGACGATCAGGCTCCTTTGAAAAAAGACTACATCATCGAACAACAGTGTGG 1564
Db 1862 ATTTCAAGGCTTTGACGAGATTACGCGGAAGAGATGTTATTGGAGAGTTGTTTGA 1921
Qy 1565 CCTTGGAGATTTTAAGACTCCTCAGAAATGTGAGAAGACCATTCTCAACCAAGGCTG 1624
Db 1922 ACTTCGCGGATTTTCGACCGCGCAGACTATTACCCGCGTGGCGGCAACAAAGGAG 1981
Qy 1625 TTTTCACAAGACAGACAAACCCAACTGTTGCTCATGTACTGAGAAGACTGT 1678
Db 1982 TCTTACAAGAACCGACAAACCCAAAGAGTGGCTCATATTCTTAGCGCGCGT 2035

RESULT 10
AR344264 1956 bp DNA linear PAT 17-AUG-2003
LOCUS AR344264
DEFINITION Sequence 1 from patent US 6582692.
ACCESSION AR344264
VERSION AR344264.1 GI:33740191
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 1956)
TITLE Podsakoff, G. and Watson, G.
JOURNAL Recombinant adeno-associated virus virions for the treatment of
FEATURES lysosomal disorders
Patent: US 6582692-A 1 24-JUN-2003;
location/Qualifiers
1..1956
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 10.2%; Score 172.4; DB 6; Length 1956;
Best Local Similarity 52.1%; Pred. No. 5e-35;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

Qy 737 TGGAACTTGAAAAAGACGAGTACACTTGGACATCGGAATCAGAACATCAGCTGGACG 796
Db 938 TGGGGCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGCTGTACCA 997
Qy 797 AGAAGAGGCTCTATCTGAACGGGAACCTGCTTTTGAAGGGCTTTGAAAAGCAGCAGG 856
Db 998 AGAGCCAGTTCCTCATCAATGGGAACCTTCTATTTCCACGGTGTCAACAAGCATGAGG 1057
Qy 857 AATTCCCGCTTCTGGGCGAGGCGACCTTTTATTCATGATGATMAAGACTTCAACTTC 916
Db 1058 ATGCGACATCCGAGGGAAGGGCTTTCAGCTGGCGCTGCTGTGAAGAACTTCAACTGC 1117
Qy 917 TGAAGTGATCAACGCGAATTCTTTCAGAGACCTTCTCACTATCCTTACAGTGAAGAGTGC 976
Db 1118 TTGGCTGGCTTGGTGCACACGCTTTCGTACCAAGCACTACCCCTATGACAGAGAGTGA 1177
Qy 977 TGGATCTTGGCCGACAGACTCGGAATCTGTGATAGCAAGCCCGCACGTTGTATCA 1036
Db 1178 TGCAGATGTGTGACCGCTATGGATGTGTGATCATGATGATGTCTCCGGCGTGGCTGG 1237
Qy 1037 CAAGGTACCA--CTACAAATCCCGAGACTCAGAAGATAGCAGAACAACATTAAGAGAA 1093
Db 1238 CGCTGCCGAGTTCTTCAACAACGTTTCTCTGATCAACCATGACAGTGTAGGAAG 1297
Qy 1094 TGATGCACAGACAAAGAACCATCCCAAGTGTGATCATGTGAGTGTGGCAAGCAACGAG 1153
Db 1298 TGGTGGTAGGGAACAAGAACACCCCGCGGTGTGATGTGTGTGTGTGGCAACGAGCTTG 1357

Qy 1154 AGTCCACCATTCCAGACCGGAGGTTTCTTCAAAAGCCCTTATGAGACTGCCAATGAAA 1213
Db 1358 GGTCCACCTAGAAATCTGTGCTGCTACTACTTGAAGATGTGATCGCTCACACCAATCCT 1417
Qy 1214 TGAATCAACACGCCCCGTGTGTATGATGTGAGCATGATGAGCAGCAGACGAGAACA 1273
Db 1418 TGGACCCCTCCCGGCTGTGACCTTTGTGAGCAACTTAATATGACG----- 1465
Qy 1274 GAGAGTGGCGCTGAAGTACTTTCAGATCGTCTGTGTGAACAGTACTACGCTGTGACA 1333
Db 1466 CAGACAGGGGGCTCCGTATGTGATGTGATCTGTGTAACAGTACTACTTGTGATC 1525
Qy 1334 TCTATCAGGGAAGATAGAGAAGACTTCAAGCTCTGAAAAAGACATAGAACGCTCT 1393
Db 1526 ACGACTACGGGACCTGAGATTGATTCAGCTGACCTGGCCACCCAGTTTGAGAACTGCT 1585
Qy 1394 ATGCAAGGCACAGAAAGCCCATCTTTGTCCAGAAATTCGGTGGCAGCGATAGTGGCA 1453
Db 1586 ATAGAAGTATCAGAAGCCCATTTATCAGAGCGAGTATGAGCAGAAACGATTGACGGGT 1645
Qy 1454 TCCACTACGATCCACTCAAAATGTTCTCCGAAGATACCAAGCAGAGCTGTTGAAAAG- 1512
Db 1646 TTCACCAAGATCCACTCTGATGTTCATGAAAGATACCAAGAAAGTCTGCTAGACAGT 1705
Qy 1513 ACGATCAGGCTCCTTTTGAAGAA-----AGACTACATCATCGGAACACAGTGTGGCCT 1567
Db 1706 ACCATCTGGGTCTGATCAAAAAGCAGAAATATGTGTTGAGAGCTCATTTGGAATT 1765
Qy 1568 TTGCAGATTTTAAGACTCCTCAGAAATGTGAGAAGCCCATTTCTCAACCAAGGCTTT 1627
Db 1766 TTGCCGATTTTCATGACTGAACAGTACCGACGAGAGTGTGGGAAATAAAAGGAGATCT 1825
Qy 1628 TCACAAGACAGACAAACCCAACTGTTGCTCATGTACTGAGA 1671
Db 1826 TCCTCGGACAGACAAACCAAAAGTGACAGCGTTCCTTTTGGCA 1869

RESULT 11
AX147652
LOCUS AX147652 1956 bp DNA linear PAT 08-JUN-2001
DEFINITION Sequence 1 from Patent WO0136603.
ACCESSION AX147652
VERSION AX147652.1 GI:14346708
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Podsakoff, G., Watson, G., Couto, L.B. and Yang, B.
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
JOURNAL 1
Recombinant adeno-associated virus virions for the treatment of
lysosomal disorders
Patent: WO 0136603-A 1 25-MAY-2001;
Avigen, Inc. (US) ; Children's Hospital Medical Center of Northern
California (US)
location/Qualifiers
1..1956
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
1..1956
/note="unnamed protein product"

CDS
/codon_start=1
/protein_id="CAC41170.1"
/db_xref="GI:14346709"
/translation="MARGSAVAAWAAAGPLLMGALGLOGMLYPOBSPSRCKEKLDEL
WSFRADFSNRRRGFEREQWYRPLWESGPTVDMPVSSFNDISQMRRLRHPVGWVYE
REVILPERWTODLRTRVYLRIQSAHSYAIWMVNGVDLLEHGGYLPFEADISNLVOYG
PLPSRLRITAIINNTLPTTLPPGTIQLYLDTSKYPKGYFVONTYFDFFNYAGLORSV
LLYTPITYIDIDITVTSVEODSGLVNYOISVKGSNLFLEVRLLDAENKVVANGTGT
OGQLKVPVSLMWPYLMHERPAYLYSLEVQLTAQTSLGVPVSDFTYLPVGIKRTVAATKS
OFLINGKPFYFHGVNKHEDADIRGKGFDMPLLVKDFNLRLWLGANAFRTSHYPAEYV
MOMCDRYGIIVVIDECBGVGLALPQFPNNVSLHHHMQVMEEVVRDKNHPAVVMMSVAN

EPASHLESAGYVLLKMWLAHTKSLDPSRPVTFVSNVNAADKGAPYVDVICLNSYSWY
HDYGHLELLIQLQLATFENWYKYOPIIOSEYGAETIYGHODPLMFTEYOKSL
EQYHGLDQKRKYVVGELIWNFADFMTEQSPTRVLGNKKGI FTROKOPKSAFLBRE
RYWKIANETRYPHSVAKSOCLENSPT"

ORIGIN	
Query Match	10.2%; Score 172.4; DB 6; Length 1956;
Best Local Similarity	52.1%; Pred. No. 5e-35;
Matches	492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;
QY	737 TGGAACTTGA AAAAGCAGTACACTTGACATCGGAATCAGAAGCATCAGCTGGACG 796
DB	938 TGGGGCCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGCTGTACCA 997
QY	797 AGAAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGGAAGCAGAG 856
DB	998 AGAGCCAGTTCCTCATCATGGGAAACCTTCTATTTCACGGGTGTCAACAAGCATGAG 1057
QY	857 AATTCCCCGTTCTGGGGCAGGGCACCTTTTATCCATTGATGATAAAGACTTCAACCTTC 916
DB	1058 ATGCGGACATCCGAGGGAAAGGCTTGCAGTGGCCGCTGCTGTGAAGGACTTCAACCTGC 1117
QY	917 TGAAGTGGATCAACCGGAATTCTTTCAGAGACCTTCATCTATCTTACAGTGAAGAGTGGC 976
DB	1118 TTCCGCTGGCTTGGTGCCTCAACGCTTTCCTGATCCAGCCACTACCCCTTATGCAAGAGAACTGA 1177
QY	977 TGGATCTTGGCCGACAGACTCGGAATCCTTGTGATAGACGAAGCCCCGACGTTGGTATCA 1036
DB	1178 TGCAGATGTGTGACCGCTATGGGATTTGTGTCATGATGAGTGTCCGGCGTGGCCCTGG 1237
QY	1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAAGATAGCAGAAGACAACATAGAAGAA 1093
DB	1238 CGCTGCCGCAAGTTCTTCAACAACGTTTCTCTGCATCACCATGCAAGTGTGAAGAAAG 1297
QY	1094 TGATGACAGACACAGAAGAACCATCCCAAGTGTGATCATGTGAGTGTGGCGAAGCAACGAG 1153
DB	1298 TGGTGCCTAGGGAACAAGAACCACCCGCGGTGCTGATGTGTCTGTGCGCAACGAGCCTG 1357
QY	1154 AGTCCAACCATCCAGACGCGGAGGGTTTCTCAAAAGCCCTTTATGAGACTGCCAATGAAA 1213
DB	1358 CGTCCCACTAGAACTGTCTGTGCTACTTGAAGATGTTGATCGCTCACACCAATCTCT 1417
QY	1214 TGGATCGAACACGCCCCGTTGTCTATGTTGAGCATGATGAGCGCACGAGACGAGAGAACAA 1273
DB	1418 TGGACCCCTCCCGGCTGTGACCTTTGTGAGCAACTCTAATCTATGAG----- 1465
QY	1274 GAGACGTGGCGCTGAAGTACTTTCGACATCGTCTGTGTGAACAGGTACTACGGCTGGTACA 1333
DB	1466 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGCTACTACTCTTGGTATC 1525
QY	1334 TCTATCAGGGAAGGATAGAAGAGGACTTCAAGCTCTGAAAAAGACATAGAAGGCTCT 1393
DB	1526 ACGACTACGGGCACTGGAGTTGATTCAGCTGCAGCTGGCCACCCAGTTTGAGAACTGGT 1585
QY	1394 ATGCAAGGCACAGAAAAGCCCATCTTTGTACAGAAATTGGTGGGACGGGATAGCTGGCA 1453
DB	1586 ATTAAGAAGTATCAGAAAGCCCATTTATTCAGAGCGAGTATGAGCAGAAAAGATTGCAAGG 1645
QY	1454 TCCAATAAGATCCACCTCAAAATGTTCTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAAAG- 1512
DB	1646 TTCAACAGGATCCACCTCTGATGTGTCACTGAAGAGTACAGAAAAGTCTGTAGAGCAGT 1705
QY	1513 ACGATCAGGCTCCTTTTGA AAAA-----AGACTACATCATCGAACAACACAGTGTGGCCCT 1567
DB	1706 ACCATCTGGGTCTGGATCAAAAAAGCAGAAAATATGTGTGAGAGCTCATTTGGAATT 1765
QY	1568 TTGCAGATTTTAAAGTCTCTCAGAAATGTGAGAAAGCCCATTTCTCAACCAAGGGTGT 1627
DB	1766 TTGCCGATTTTCATGATGAACAGTCAACGACGAGAGTGTCTGGGGAATAAAAAGGGATCT 1825
QY	1628 TCACAAGAGACAGAACCCAAACTCGTGTGCTCATGTACTGAGA 1671
DB	1826 TCACTCGGCAGAGACAACCAAAAAGTGCAGCGTTCCTTTGGCA 1869

RESULT 12
CQ847932
LOCUS CQ847932 2191 bp DNA linear PAT 19-AUG-2004
DEFINITION Sequence 237 from Patent WO2004063362.
ACCESSION CQ847932
VERSION CQ847932.1 GI:51469520
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS Glover,D., Bell,G., Frenz,L. and Midgley,C.
TITLE Cell cycle progression proteins
JOURNAL Patent: WO 2004063362-A 237 29-JUL-2004;
Cyclacel Limited (GB)
FEATURES
source location/Qualifiers
1..2191
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="M15182"

ORIGIN	
Query Match	10.2%; Score 172.4; DB 6; Length 2191;
Best Local Similarity	52.1%; Pred. No. 5e-35;
Matches	492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;
QY	737 TGGAACTTGA AAAAGCAGTACACTTGACATCGGAATCAGAAGCATCAGCTGGACG 796
DB	964 TGGGGCCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGCTGTACCA 1023
QY	797 AGAAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGGAAGCAGAG 856
DB	1024 AGAGCCAGTTCCTCATCATAGGGAACCTTTCTATTTCACGGGTGTCAACAAGCATGAG 1083
QY	857 AATTCCCCGTTCTGGGGCAGGGCACCTTTTATCCATTGATGATTAAGACTTCAACCTTC 916
DB	1084 ATGCGGACATCCGAGGGAAGGGCTTGCAGTGGCCGCTGTGTGAAGGACTTCAACCTGC 1143
QY	917 TGAAGTGAATCAACGCGAATTTCTTACAGACCTCTCACTATCTTACAGTGAAGAGTGGC 976
DB	1144 TTGCTGGCTTGGTGTGCCAAGCGCTTTCCTGATCCAGCCACTACCCCTATGCAAGGAAGTGA 1203
QY	977 TGGATCTTGGCCGACAGACTCGGAATCCTTGTGATAGACGAAGCCCCGACGTTGGTATCA 1036
DB	1204 TGCAGATGTGTGACCGCTATGGGATTTGTGTGATCATGATGAGTGTCCCGGCTGGCCCTGG 1263
QY	1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAAGATAGCAGAAGACAACATAGAAGAA 1093
DB	1264 CGCTGCCGAGTCTTCAACAACGTTTCTGTGATCACCATGCAAGTGCATGGAAGAAAG 1323
QY	1094 TGATGACAGACACAGAAGAACCATCCCAAGTGTGATCATGTGAGAGTGTGGCGAAGCAACGAG 1153
DB	1324 TGGTGCCTAGGGAACAAGAACCACCCGCGGTGCTGATGTGTCTGTGCGCAACGAGCCTG 1383
QY	1154 AGTCCAACCATCCAGACGCGGAGGGTTTCTCAAAAGCCCTTTATGAGACTGCCAATGAAA 1213
DB	1384 CGTCCACCTAGAAATCTGTGCTACTACTTGAAGATGTTGATGCTCACACCAATCTCT 1443
QY	1214 TGGATGAAACAGCCCCGTTGTCTATGTTGAGCATGATGAGCGCACGAGCAGAGAAACAA 1273
DB	1444 TGGACCCCTCCCGGCTGTGAACCTTTGTGAGCAACTCTAATCTATGAG----- 1491
QY	1274 GAGACGTGGCGCTGAAGTACTTGCACATCGTCTGTGTGAACAGTACTACGGCTGGTACA 1333
DB	1492 CAGACAAGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGCTACTACTCTTGGTATC 1551
QY	1334 TCTATCAGGGAAGGATAGAAGAGGACTTCAAGCTCTGAAAAAGACATAGAAGAGCTCT 1393
DB	1552 ACGACTACGGGCACTGAGTTGATTCAGCTGCAGCTGGCCACCGAGTTGAGAACTGCT 1611

QY

1394

ATGCAAGCACAGAAAGCCATCTTGTCA

AGAATTCGGTCCGACCGATAGCTGGCA

1453

Db

1612

ATAAGAAGTATCAGAAGCCATTATTCA

GAGCGAGTATGAGACAGAAACGATTGCAGGGT

1671

QY

1454

TCCACTACGATCCACCTCAATGTTCTCC

GGAAGAGTACCACAGACAGAGCTCGTTGAAAG-

1512

Db

1672

TTCAACGAGATCCACCTCTGATGTTCAC

TGAAGATACCAGAAAAGTCTGCTAGAGCACT

1731

QY

1513

ACGATCAGGCTCCTTTGAAAAA-----

AGACTACATCATCGGAACACAGTGTGGGCT

1567

Db

1732

ACCATCTGGGTCTGGAATCAAAAAACG

AGAAAATATGTGTTGGAGAGCTCATTTGGAATT

1791

QY

1568

TTGCAGATTTTAAGACTCTCAGATGTGA

GAGAACCCATTCTCAACCAAGGCTTTT

1627

Db

1792

TTGCCGATTTTCATGACTGAACAGTCAC

CGACGAGAGTGTGGGAATRAAAAGGGGATCT

1851

QY

1628

TCACAGAGACAGACACCCCAACTCGTT

GCTCATGTACTGAGA

1671

Db

1852

TCACCTCGCAGAGACAAACCAAAAGTG

CAGCGTTCCTTTTGC

1895

RESULT 13

HUMGLCB

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

2191 bp

mRNA

linear

PRI 08-NOV-1994

Human beta-glucuronidase mRNA, complete cds.

M15182

M15182.1

GI:183232

alternative splicing; beta-D-glucuronoside glucuronosohydrolase; beta-glucuronidase; glucuronidase; glucuronohydrolase; hydrolase.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.

1 (bases 1 to 2191)

Oshima,A., Kyle,J.W., Miller,R.D., Hofmann,J.W., Powell,P.P., Grubb,J.H., Sly,W.S., Tropak,M., Guise,K.S. and Gravel,R.A.

Cloning, sequencing, and expression of cDNA for human beta-glucuronidase

Proc. Natl. Acad. Sci. U.S.A. 84 (3), 685-689 (1987)

87118233

3468507

Original source text: Human fibroblasts cDNA clone pHUGF and placenta cDNA clones pHUGP13 and pHUGP15.

Draft entry and copy of computer-readable sequence of [1] kindly provided by R.D.Miller, 08-MAY-1987.

Four potential N-linked glycosylation sites are located at bases 543-551, 840-848, 1284-1292 and 1917-1925. Potential poly-A signals were found 25 and 48 bp upstream of the poly-A site at nucleotides 2139-2144 and 162-2167 respectively.

Two types of mRNA exit due to alternative splicing. The short form, lacking bases 939-1091 below, encodes a protein lacking beta-glucuronidase activity.

Location/Qualifiers

1..2191

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/map="7q22"

1..2191

/gene="GUSB"

<1..2191

/gene="GUSB"

/product="glucuronidase-b mRNA"

27..1982

/gene="GUSB"

/note="beta-glucuronidase precursor (EC 3.2.1.31)"

/codon_start=1

/protein_id="AAA52561.1"

/db_xref="GI:183233"

/db_xref="GDB:G00-120-025"

/translation="MARGSAVVAAMALGPLWGALGLOGMLYPOESPSECKELDGLWSFRADFSDNRRRGFEQWYRRPLWESGPTVDMVPSPSFNDISQDMRLRHFGVWVWE

REVILPERWTODLRTRVLRIGSAHSYAIVWNGVDLTLEHGGYLPEADISNLVQVG

PLPSRLRITIAINNTLPTTLRPGTIOYLTDTSKYPKGYFVQNTYFDFFNYAGLQRSV

LYLTPTTYIDDIITWTSVEQDLVNYQISVKGSNLKLEVLRLDAENKRVANVTGT

OGQLKVPGLVSLWPLYLMEHRDAYLYSLEVQLTAQTSIGLPSVDFYTLPGVIRTVATKS

QFLINGKPEYFHVGNKEDADI RKGCFDPLLVKDFNLRLGANAFTSHYPYAEV

MQMCDRYGIIVIDECPVGLALPQFENNVS LHHMQWMEVVRDKHNAVVMMSVAN

EPASHLESAGYLLKMTIAHTKSLDSPRPTFVNSNSYAADKGA PYVDVCLNSYSWY

HDYGHLELIQLQLATQFENWYKQKPIIQSEYGAETIAGFHQDPPLMETBEYQKSL

EQYHLGLDQKRKYVWGBLIWNFADFMTEQSPTRVLGNKKGI FTRORQPSAAFLBRE

RYWKIANBTRYPHSVAKSOCLENSPT"

27..92

/gene="GUSB"

/note="beta-glucuronidase signal peptide"

93..1979

/gene="GUSB"

/product="beta-glucuronidase mature peptide"

ORIGIN

33 bp upstream of SmaI site.

Query Match

10.2%; Score 172.4; DB 9; Length 2191;

Best Local Similarity 52.1%; Pred. No. 5e-35;

Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

QY

737

TGGAATTGAAAAAGACGAGTACACTCTG

ACATCGGAATCAGAACGATCAGCTGGACG

796

Db

964

TGGGGCCTGTGTGACTTCTACACACTCC

CTGTGGGATCCGCACTGTGCTGCACCA

1023

QY

797

AGAGAGGCTCTATCTGAACGGGAAACCT

GTCTTTTGAAGGCTTTGAAAGACGAGG

856

Db

1024

AGAGCCAGTTCCTCATCATGAGGAAAC

CTTCTATTTCCACGGTGTCAACAAGCAT

1083

QY

857

AATTCGCCGTTCTGGGGCAGGGACCTTT

TATCCATGTGATTAAGAAGCTTCAACTTC

916

Db

1084

ATGCGGACATCCGAGGGAAGGCTTCGAC

TGGCGCTGCTGTGAAGACTTCAACCTGC

1143

QY

917

TGAAGTGATCAACGGCAATCTTTCAGAC

CACTCTCACTATCTTACAGTGAAGGTGC

976

Db

1144

TTGCGTGGCTTGGTGCCAAAGCTTCCGT

ACCAAGCACTACCTTATGAGAGGAAGTGA

1203

QY

977

TGATCTTGGCCAGACAGACTCGGAATCT

TGTGATAGACGAAGCCCGCACGTTGTATCA

1036

Db

1204

TGCAGATGTGTGACCGCTATGGGATTGT

GTGATGATGATGTGCCGCTGGCTGG

1263

QY

1037

CAAGGTACCA---CTACAATCCCGAGACT

CAAGAGATAGCAGAGACAAACATAGAAGAA

1093

Db

1264

CGCTGCCGCAAGTTCTTCAACAACGTTT

CTCTGCATCAACCAATGCAAGGTGATGGAAG

1323

QY

1094

TGATCGACAGACACAAGAACCATCCAGT

GTGATCATGTGAGTGTGGGAAACCAACAG

1153

Db

1324

TGGTGCCTAGGGACAAAGAACCCCGGGT

CGTGATGTGTGTGTGGCCAAACGACCTG

1383

QY

1154

AGTCCAACCATCCAGACGGCGAGGGTTT

CTTCAAAAGCCCTTTATGAGACTGCCAATGA

1213

Db

1384

CGTCCACCTAGAAATCTGCTGGCTACTAC

TGAAAGATGTGATGCTCACACCAAAATCCT

1443

QY

1214

TGGATCGAACACCGCCCGTGTGTGATGT

GAGCATGATGAGCGCACAGACAGAGAACAA

1273

Db

1444

TGGAACCTCTCCGGCCTGTGACCTTTGT

GAGCAACTTAATATGAG-----

1491

QY

1274

GAGACGTGGCGCTGAAGTACTTGCATCT

GTGTGTGAACAGGTACTACGGCTGGTACA

1333

Db

1492

CAGACAAAGGGGCTCCGTATGTGATGTG

ATCTGTTGAACAGCTACTACTTGGTATC

1551

QY

1334

TCATTCAGGGAGGATAGAGAAAGACTTCA

AGCTCTGAAAAAGACATAGAAAGACTCT

1393

Db

1552

ACGACTACGGGCACTGAGTTGATTCAGCT

GAGCTGGCCACCAAGTTGAGAACTGGT

1611

QY

1394

ATGCAAGGCACAGAAAGCCCATCTTTGT

TCACAGAAATTCGGTGGGAGCGGATAGCTGGCA

1453

Db

1612

ATAAGAAGTATCAGAAGCCCAITTAATTC

AGAGCGAGTATGAGCAGAAAAGATTGCAAGGT

1671

QY

1454

TCCACTACGATCAACTCAATGTCTCCGAA

GAGTACCAAGCAGAGCTCGTTGAAAAAG-

1512

Db

1672

TTCAACAGGATCACTCTGATGTTCAC

TGAAGATACCAAGAAAGTCTGTGTAAGCAG

1731

ACCESSION AX827491
 VERSION AX827491.1 GI:39837679
 KEYWORDS
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1
 AUTHORS Boese, F., Suter-Dick, L. and Wolf, D.
 TITLE Methods for the toxicity prediction of a compound
 JOURNAL Patent: EP 1344834-A 225 17-SEP-2003;
 F. HOFFMANN-LA ROCHE AG (CH)
 FEATURES Location/Qualifiers
 source 1..2472
 /organism="Rattus norvegicus"
 /mol_type="unassigned DNA"
 /db_xref="taxon:10116"
 ORIGIN

Query Match 9.9%; Score 166.8; DB 6; Length 2472;
 Best Local Similarity 51.3%; Pred. No. 1.6e-33;
 Matches 477; Conservative 0; Mismatches 432; Indels 21; Gaps 3;

QY 751 GACGAGTACACTGTGACATCGAATCAGAAGATCAGCTGGACGAGAAGAGGCTCTAT 810
 DB 946 GACTTCTATACCTCCCTGTCGGGATTCGAACAGTGGCTGTACAAAGACAAAGTTCCCTC 1005
 QY 811 CTGAACGGGAAACCTGTCTTTTGAAGGCTTTGGAAGACGAGGAATCCCCCTCTG 870
 DB 1006 ATAAATGGGAAGCCTTTCTACTTCAAGGCGTCAACAAGCATGAGATTCAAGATATCCGA 1065
 QY 871 GGGCAGGGCACCTTTATCCATTGATGATAAAGACTTCAACCTTCTGAAGTGAATCAAC 930
 DB 1066 GGGAGAGGCTTGAAGCTGAGCTGCTGATGATAAAGATTCAACCTCCTCGTTGGCTCGGG 1125
 QY 931 GCGAATTTCTTGAAGACCTCTCACTATCTTACAGTGAAGATGGCTGATCTTCCCGAC 990
 DB 1126 GCAAAATTCCTTTCGTAACGACCACTATCCCTACTCGGAGAGGATCTTCAAGCTCTGTGAC 1185
 QY 991 AGACTCGGAATCCTTGTGATAGAGAACCCCGCACTTGTGATCAAGGTACCA---C 1047
 DB 1186 CGATATGGAATGTGTCATGATGATGTCCCGGTGTGGCATCTGTGCTGCCCAAGAT 1245
 QY 1048 TACAATCCCGAGACTCAGAAGATGACAGAACAATAGAAGATGATCGACAGACAC 1107
 DB 1246 TTGGCAACGTGTCTTCGCGCACCACTAGAGGTGATGACGAGCTGTGCGCAGGGAC 1305
 QY 1108 AAGAACCATCCCACTGTGATCATGTGAGTGTGGGAAAGAACCAAGATCCAAACCATCCA 1167
 DB 1306 AAAAATCACCCCTGCGGCTGTGATGTGTCTGTGCGCAATGAGCCTGTCTTCTTGAAA 1365
 QY 1168 GACCGCGAGGGTTTCTTCAAAGCCCTTATGAGACTGCCAATGAATGATCGAACACGC 1227
 DB 1366 CCTGCGGATATTAATTCAAGACCTGATCGCCACACCAAAAGCCCTGAGCCCAACCCGT 1425
 QY 1228 CCCGTTGTCATGTGAGCATGATGAGACGACCAAGAGAGAGAACAGACCTGGCGCTG 1287
 DB 1426 CCGTGAACCTTGTGACCAATA-----CCAGATATGACGACATGGGGGCC 1473
 QY 1288 AAGTACTTGCACATCGTCTGTGTGAACAGGTACTACGGCTGTGATCATCTATCAGGGAAAG 1347
 DB 1474 CCGTACGTGACGTGATTTGTGTGAACAGTTACTTATCTGTGATCATGACTACGGGCAT 1533
 QY 1348 ATAGAAGAAAGACTTCAAGCTCTGAAAAAGACATAGAAGAGCTCTATGCAAGGACACAGA 1407
 DB 1534 CTGAGGTGATTCAGCTGCAAGCTGACTAGCCAGTTTGAAGAACTGTATTAAGATGTAACAG 1593
 QY 1408 AAGCCCATCTTTGTCAACGAATTCGGTGGAGCGGATAGCTGGCATCCACTACGATCCA 1467
 DB 1594 AAGCCAATTATCCAGAGCGAGTATGAGACGACGCCGTCTCGGGGCTTCATGAGGATCCA 1653
 QY 1468 CCTCAATATGTTCTCGAAGAGTACCAAGCAGAGCTGTTGAAAAAGAGATCAGGCTCCTT 1527

DB 1654 CCTGCATGTTTCACTGAGAGTACCAAGACAGCTCTCCCTGAGAAATTATCATTTGATCCTG 1713
 QY 1528 TTGAA-----AAAGACTATCATTCGGAACACACGTGTGGCCCTTTCAGATTTTAAG 1581
 DB 1714 GATGAGAAACGAAAAAATATGTATTCGGAAGCTCATCTGGAATTTTGTGACTTCATG 1773
 QY 1582 ACTCCTCAGAATGTGAGAACCCATTTCTAACCAACAAGGTTTTCACAAGAGACAGA 1641
 DB 1774 ACCAAACAGTACCACTGAGAGTAAACAGAAACAGAAAGGGGATCTTCACTCGACAGAGA 1833
 QY 1642 CAACCCAAACTCGTTGCTCATGTACTGAGA 1671
 DB 1834 AACCCCAAGATGGCAGCCCTTCATTTTGGCA 1863

Search completed: September 2, 2005, 00:30:40
 Job time : 7498 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2005, 19:07:51 ; Search time 975 Seconds
(without alignments)
10254.813 Million cell updates/sec

Title: US-09-936-759-14
Perfect score: 1689
Sequence: 1 atggttaagaccgcaacgaaa.....gaagactgtgagtgaggtc 1689

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*

1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1688	99.9	1689	3	AAA07937	Aaa07937 Thermotoga
2	1688	99.9	1692	13	ADS48252	Ads48252 Bacterial
3	174	10.3	2050	13	ACN43424	Acn43424 Human dia
4	174	10.3	2082	13	ACN43423	Acn43423 Human dia
5	174	10.3	2128	4	ABL04147	Abi04147 Drosophila
6	174	10.3	2169	8	AAD50922	Aad50922 L. mexicana
7	174	10.3	2169	12	ADF47502	Adf47502 Human bet
8	174	10.3	2169	12	ADJ58608	Adj58608 Fusion DN
9	172.4	10.2	1956	4	AAD06386	Aad06386 Human bet
10	172.4	10.2	2191	6	ABK83932	Abk83932 Human cdN
11	172.4	10.2	2191	10	ADD29602	Add29602 Human tum
12	172.4	10.2	2191	10	ADG89395	Adg89395 Cancer de
13	172.4	10.2	2191	12	ACF57506	Acf57506 Human bet
14	172.4	10.2	2191	12	ADJ62864	Adj62864 Human bet
15	172.4	10.2	2191	12	ADP10326	Adp10326 Reference
16	172.4	10.2	2191	13	ADQ89807	Adq89807 Antagonis
17	172.4	10.2	3525	4	AAF57574	Aaf57574 Mutant he
18	172.4	10.2	7328	4	AAF57573	Aaf57573 Heavy cha
19	170.8	10.1	3314	2	AAQ58896	Aaq58896 Humanised
20	166.8	9.9	2472	10	ADB58226	Adb58226 Toxicity-

21	166.8	9.9	2472	10	ADB52752	Adb52752 Primary r
22	159.6	9.4	1947	4	AAD06387	Aad06387 Murine be
23	145.8	8.6	1929	13	ACN43422	Acn43422 Human dia
24	145.8	8.6	2109	10	ADG76418	Adg76418 Human inc
25	142.2	8.4	23078	4	AAS59508	Aas59508 Propionib
26	142.2	8.4	23078	8	ACF64437	Acf64437 Propionib
27	141.4	8.4	1887	2	AAX23826	Aax23826 Bacillus
28	133	7.9	1820	3	AAA07958	Aaa07958 Staphyloc
29	132.6	7.9	1812	3	AAA07939	Aaa07939 Salmonella
30	130.4	7.7	1888	3	AAA07938	Aaa07938 Codon-opt
31	114	6.7	1814	10	ADC01686	Adc01686 Enterohae
32	114	6.7	3451	8	ABT16609	Abt16609 Artificia
33	114	6.7	3451	10	ACC44711	Acc44711 Plasmid p
34	114	6.7	14627	8	ABT16610	Abt16610 Artificia
35	114	6.7	14627	10	ACC44712	Acc44712 Plasmid p
36	110.8	6.6	1809	6	ABK87080	Abk87080 uida DNA
37	110.8	6.6	1812	4	AAH22503	Aah22503 PCR gener
38	110.8	6.6	1812	6	AAD29543	Aad29543 uida DNA
39	110.8	6.6	1812	6	ABL61371	Abi61371 GUS DNA #
40	110.8	6.6	1812	8	ABT16606	Abt16606 Artificia
41	110.8	6.6	1812	10	ACC44709	Acc44709 E. coli b
42	110.8	6.6	1812	13	ADS46099	Ads46099 Bacterial
43	110.8	6.6	2001	4	AAI19830	Aai19830 Beta-gluc
44	110.8	6.6	2001	6	ABK15667	Abk15667 Glutathio
45	110.8	6.6	2001	13	ADR16848	Adr16848 Arabidops

ALIGNMENTS

RESULT 1	
AAA07937	
ID AAA07937 standard; DNA; 1689 BP.	
XX	
AC AAA07937;	
XX	
DT 26-JAN-2001 (first entry)	
XX	
DB Thermotoga maritima beta-glucuronidase gene.	
XX	
KW Microbial; beta-glucuronidase; GUS; Enterobacter; Salmonella;	
KW Pseudomonas; Staphylococcus; Thermotoga; transgenic plant; Bioindicator;	
KW transgenic insect; marker; glucuronide detoxification; ds.	
XX	
OS Thermotoga maritima.	
XX	
PN WO200055333-A1.	
XX	
PD 21-SEP-2000.	
XX	
PF 16-MAR-2000; 2000WO-US007107.	
XX	
PR 17-MAR-1999; 99US-00270957.	
XX	
PA (CAMB-) CAMBIA BIOSYSTEMS LLC.	
XX	
PI Jefferson RA, Mayer JE;	
XX	
DR WPI; 2000-647075/62.	
DR P-PSDB; AAB28406.	
XX	
PT Novel microbial beta-glucuronidase genes and gene products used as	
PT reporter/effector molecule, as diagnostic tool, in positive selection, to	
PT target molecules to specific cells and to detect and track linked genes.	
XX	
PS Claim 2; Fig 4; 116pp; English.	
XX	
CC The present sequence encodes a microbial beta-glucuronidase (GUS)	
CC protein. GUS genes were obtained from six different genera:	
CC Enterobacter/Salmonella, Pseudomonas, Staphylococcus and	
CC Thermotoga. Microbial GUS can be used as a reporter/effector molecule for	
CC transgenic constructions and in in vitro diagnostic applications. It may	
CC also be used to generate sentinel plants that serve as bioindicators of	

CC environmental status. It may be used to generate transgenic insects for
CC tracking insect populations or to facilitate the development of a
CC bioassay for compounds that affect molecules critical for insect
CC development (e.g. juvenile hormone). Secreted GUS may also serve as a
CC marker for beneficial fungi destined for release into the environment. In
CC animal systems, secreted GUS may be used to achieve extracellular
CC detoxification of glucuronides (e.g. toxin glucuronide) and to examine
CC conjugation patterns of glucuronides. Microbial GUS may also be used in
CC traditional medical diagnostic assays, for drug testing, pharmacokinetic
CC studies, bioavailability studies, diagnosis of diseases and syndromes,
CC following progression of disease or its response to therapy. Microbial
CC GUS has increased thermal stability, high turnover number and enzymatic
CC activity. It is highly specific for the substrate and water soluble, and
CC the substrates are stable

XX Sequence 1689 BP; 526 A; 374 C; 430 G; 358 T; 0 U; 1 Other;

Query Match 99.9%; Score 1688; DB 3; Length 1689;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1689; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTAAGACCGCAACGAAGAAGATTATCTTATCTTGAATGAGTTTGAAT 60
DB 1 ATGCTAAGACCGCAACGAAGAAGATTATCTTATCTTGAATGAGTTTGAAT 60
QY 61 CTTGAAGTAACCAAGACAGACCAATCCGCTCTGGAAGCTGGAATGAGCACTAC 120
DB 61 CTTGAAGTAACCAAGACAGACCAATCCGCTCTGGAAGCTGGAATGAGCACTAC 120
QY 121 CAGGATCTGTGCTACGAAGAAGACCCCTTCACTCAAAAACCACTTCTACGTTCCGAAG 180
DB 121 CAGGATCTGTGCTACGAAGAAGACCCCTTCACTCAAAAACCACTTCTACGTTCCGAAG 180
QY 181 NAACTTTCACAAAACACATCAGACTTTACTTGTGCGGTGAACACGCACTGCGAGTC 240
DB 181 NAACTTTCACAAAACACATCAGACTTTACTTGTGCGGTGAACACGCACTGCGAGTC 240
QY 241 TTCTTCAACGAGAGAAAGTGGAGAAATCACAATTGATACCTTCCCTCGAAGTAGAT 300
DB 241 TTCTTCAACGAGAGAAAGTGGAGAAATCACAATTGATACCTTCCCTCGAAGTAGAT 300
QY 301 GTGACGGGGAAAGTGAATCCGAGAGAAAGCACTCAGGGTGTGTGAGAACAGATTG 360
DB 301 GTGACGGGGAAAGTGAATCCGAGAGAAAGCACTCAGGGTGTGTGAGAACAGATTG 360
QY 361 AAAGTGGAGAGATTCCCTCGAAGTTCAGACAGCGGCACTCACACCGTGGATTTTT 420
DB 361 AAAGTGGAGAGATTCCCTCGAAGTTCAGACAGCGGCACTCACACCGTGGATTTTT 420
QY 421 GGAAGTTTCCACCTGCAAACTTCGACTTCTTCCCTACGGTGAATCATAGGCTGTT 480
DB 421 GGAAGTTTCCACCTGCAAACTTCGACTTCTTCCCTACGGTGAATCATAGGCTGTT 480
QY 481 CTGATAGAGTTCACAGACCACGCGAGATACTCGACATCTGGGTGACACAGAGTAGTCT 540
DB 481 CTGATAGAGTTCACAGACCACGCGAGATACTCGACATCTGGGTGACACAGAGTAGTCT 540
QY 541 GAACCGGAGAAAGAACTTGGAAAAGTGAAGTAGAAGTCTCAGAAAGACGGTG 600
DB 541 GAACCGGAGAAAGAACTTGGAAAAGTGAAGTAGAAGTCTCAGAAAGACGGTG 600
QY 601 GGACAGAGATGACGATCAAACTTGGAGAGAGAAAAAGATTAGAACATCCACACAG 660
DB 601 GGACAGAGATGACGATCAAACTTGGAGAGAGAAAAAGATTAGAACATCCACACAG 660
QY 661 TTGCTCGAAGGGGAGTTTCCTCGAAAACGCCAGGTTCTGAGCCTCGAAGATCATAT 720
DB 661 TTGCTCGAAGGGGAGTTTCCTCGAAAACGCCAGGTTCTGAGCCTCGAAGATCATAT 720
QY 721 CTTTATCTCTCAAGGTGGAAGTGAAGAAAGACGAGTAGACTCTGAGCATCGGAATCAGA 780
DB 721 CTTTATCTCTCAAGGTGGAAGTGAAGAAAGACGAGTAGACTCTGAGCATCGGAATCAGA 780

QY 781 ACGATCAGCTGGGACGAGAAAGAGGCTCTATCTGAAACGGGAAACCTGTCTTTTGAAGGC 840
DB 781 ACGATCAGCTGGGACGAGAAAGAGGCTCTATCTGAAACGGGAAACCTGTCTTTTGAAGGC 840
QY 841 TTTGAAAGACGAGGAATTCCTGTTCTGGGGCAGGGCACCTTTATCCATTGATGATA 900
DB 841 TTTGAAAGACGAGGAATTCCTGTTCTGGGGCAGGGCACCTTTATCCATTGATGATA 900
QY 901 AAAGACTTCAACCTTCTGAAAGTGAATCAACCGGAATCTTTTCAAGACCTCTCATCT 960
DB 901 AAAGACTTCAACCTTCTGAAAGTGAATCAACCGGAATCTTTTCAAGACCTCTCATCT 960
QY 961 TACAGTGAAGAGTGGCTGGATCTTGGCGACGACTCGGAATCTTGTATAGCGAAGCC 1020
DB 961 TACAGTGAAGAGTGGCTGGATCTTGGCGACGACTCGGAATCTTGTATAGCGAAGCC 1020
QY 1021 CCGCAGCTTGGTATCACAAGGTACCACTACAAATCCCGAGACTCAGAAATAGCAGAAAG 1080
DB 1021 CCGCAGCTTGGTATCACAAGGTACCACTACAAATCCCGAGACTCAGAAATAGCAGAAAG 1080
QY 1081 AACATAGAAGAAATGATCGACAGACACAAAGAACATCCAGTGTGATCATGTGAGTGTG 1140
DB 1081 AACATAGAAGAAATGATCGACAGACACAAAGAACATCCAGTGTGATCATGTGAGTGTG 1140
QY 1141 GCGAAACGAACAGAGTCCAAACCATCCAGCGGGAGGTTTCTTCAAGCCCTTTATGAG 1200
DB 1141 GCGAAACGAACAGAGTCCAAACCATCCAGCGGGAGGTTTCTTCAAGCCCTTTATGAG 1200
QY 1201 ACTGCCAATGAATGATCGAACACGCCCCGTTGTCTATGTGAGCATGATGAGCGACCA 1260
DB 1201 ACTGCCAATGAATGATCGAACACGCCCCGTTGTCTATGTGAGCATGATGAGCGACCA 1260
QY 1261 GACGAGAGAACAGAGACGTGGCGCTGAAGTACTTCGACATCTGTGTGAACAGGTAC 1320
DB 1261 GACGAGAGAACAGAGACGTGGCGCTGAAGTACTTCGACATCTGTGTGAACAGGTAC 1320
QY 1321 TACGGCTGTGATCTATCAGGGAAGGATAGAAGAGACTTCAAGCTCTGAAAAAGAC 1380
DB 1321 TACGGCTGTGATCTATCAGGGAAGGATAGAAGAGACTTCAAGCTCTGAAAAAGAC 1380
QY 1381 ATAGAAGAGCTCTATGCAAGGCAAGAAAGCCATCTTGTCAAGAAATTCGTCGGGAC 1440
DB 1381 ATAGAAGAGCTCTATGCAAGGCAAGAAAGCCATCTTGTCAAGAAATTCGTCGGGAC 1440
QY 1441 GCGATAGCTGGCATCCACTAGATCCACTCAATGTTCTCGAAGAGTACCAAGCAGAG 1500
DB 1441 GCGATAGCTGGCATCCACTAGATCCACTCAATGTTCTCGAAGAGTACCAAGCAGAG 1500
QY 1501 CTCGTTGAAAAAGACGATCAGGCTCTTTGAAAAAAGACTACATCATCGAACACAGTG 1560
DB 1501 CTCGTTGAAAAAGACGATCAGGCTCTTTGAAAAAAGACTACATCATCGAACACAGTG 1560
QY 1561 TGGGCTTTGAGATTTTAAGACTCCTCAGAAATGTGAGAAAGCCATTCTCAACCAAG 1620
DB 1561 TGGGCTTTGAGATTTTAAGACTCCTCAGAAATGTGAGAAAGCCATTCTCAACCAAG 1620
QY 1621 GGTGTTTTCACAAGAGACAGCAAACTCGTTGCTCATGTACTGAGAAAGCTGTGG 1680
DB 1621 GGTGTTTTCACAAGAGACAGCAAACTCGTTGCTCATGTACTGAGAAAGCTGTGG 1680
QY 1681 AGTGAGGTT 1689
DB 1681 AGTGAGGTT 1689

RESULT 2
ADS48252
ID ADS48252 standard; cDNA; 1692 BP.
XX
AC ADS48252;
XX
DT 02-DEC-2004 (first entry)
XX

DE Bacterial polynucleotide #2995.
 XX
 KW Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polynucleotide; gene; ss.
 XX
 OS Bacteria.
 XX
 PN US2003233675-A1.
 XX
 PD 18-DEC-2003.
 XX
 PF 20-FEB-2003; 2003US-00369493.
 XX
 PR 21-FEB-2002; 2002US-0360039P.
 XX
 PA (CAOY/) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.
 XX
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 XX
 DR WPI; 2004-061375/06.
 XX
 PT New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.
 XX
 PS Claim 1; SEQ ID NO 26682; 122bp; English.
 XX
 CC The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polynucleotide used in
 CC the scope of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.
 CC
 SQ Sequence 1692 BP; 527 A; 374 C; 432 G; 359 T; 0 U; 0 Other;
 Query Match 99.9%; Score 1688; DB 13; Length 1692;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1688; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ATGCTAAGACCGCAAGCAAGAGAGATTATCTTATCTTGAATGAGTTTGAAT 60
 Db 1 ATGCTAAGACCGCAAGCAAGAGAGATTATCTTATCTTGAATGAGTTTGAAT 60
 QY 61 CTGTAAGTAACCAAGCAAGACAGACCAATCGCGTTCTGGAAGCTGGAATGAGCAGTAC 120
 Db 1 CTGTAAGTAACCAAGCAAGACAGACCAATCGCGTTCTGGAAGCTGGAATGAGCAGTAC 120

Db 61 CTGTAAGTAACCAAGCAAGACAGACCAATCGCGTTCTGGAAGCTGGAATGAGCAGTAC 120
 QY 121 CAGGATCTGTGCTACGAAGAGACCCCTTCACTACCAAAAACCACTTCTACGTTCCGAAG 180
 Db 121 CAGGATCTGTGCTACGAAGAGACCCCTTCACTACCAAAAACCACTTCTACGTTCCGAAG 180
 QY 181 NAACCTTACAAAAACATCAGACTTTTACTTGTGCGGTGAACACGGAAGTCCGAGGTC 240
 Db 181 GAACCTTACAAAAACATCAGACTTTTACTTGTGCGGTGAACACGGAAGTCCGAGGTC 240
 QY 241 TTCTCAACGAGAGAAAGTGGAGAGAAATCAATGAATACCTTCCCTTGAAGTAGAT 300
 Db 241 TTCTCAACGAGAGAAAGTGGAGAGAAATCAATGAATACCTTCCCTTGAAGTAGAT 300
 QY 301 GTGACGGGGAAGTGAATCCGAGAGAACGAACCTCAGGGTGTGTTGAGAACAGATTG 360
 Db 301 GTGACGGGGAAGTGAATCCGAGAGAACGAACCTCAGGGTGTGTTGAGAACAGATTG 360
 QY 361 AAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACCGTGAGATTTT 420
 Db 361 AAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACCGTGAGATTTT 420
 QY 421 GGAAGTTTTCCACCTGCAAACTTGCATTCTTCCCTACGGTGAATCATAGGCGTGT 480
 Db 421 GGAAGTTTTCCACCTGCAAACTTGCATTCTTCCCTACGGTGAATCATAGGCGTGT 480
 QY 481 CTGATAGAGTTCAAGACCAACGCGAGGATCTCGACATCTGGGTGACACGAGTAGTCT 540
 Db 481 CTGATAGAGTTCAAGACCAACGCGAGGATCTCGACATCTGGGTGACACGAGTAGTCT 540
 QY 541 GAACCGGAGAGAAACTTGAAGAGTGAAGTGAAGATAGAAGTCTCAGAGAGCGGTG 600
 Db 541 GAACCGGAGAGAAACTTGAAGAGTGAAGTGAAGATAGAAGTCTCAGAGAGCGGTG 600
 QY 601 GGAAGAGAGATGAGATCAAACTTGAAGAGAGAGAGAGAGAGAGAGAGATCAACAGAG 660
 Db 601 GGAAGAGAGATGAGATCAAACTTGAAGAGAGAGAGAGAGAGAGAGATCAACAGAG 660
 QY 661 TTGCTCGAAGGGGAGTTTCATCTCCGCAAAACGCCAGCTTCTGAGCTCGAAGATCCATAT 720
 Db 661 TTGCTCGAAGGGGAGTTTCATCTCCGCAAAACGCCAGCTTCTGAGCTCGAAGATCCATAT 720
 QY 721 CTTATCTCTCAAGGTGGAATTTGAAAAAGACAGATACACTCTGACATCGGAATCAGA 780
 Db 721 CTTATCTCTCAAGGTGGAATTTGAAAAAGACAGATACACTCTGACATCGGAATCAGA 780
 QY 781 ACGATCAGCTGGGACGAGAGAGAGCTCTATCTGAACGGGAACCTGCTTTTGAAGGGC 840
 Db 781 ACGATCAGCTGGGACGAGAGAGAGCTCTATCTGAACGGGAACCTGCTTTTGAAGGGC 840
 QY 841 TTTGAAAAACGAGGAATTTCCCGTTCTGGGGCAGGGCACCTTTTATCCATTGATGATA 900
 Db 841 TTTGAAAAACGAGGAATTTCCCGTTCTGGGGCAGGGCACCTTTTATCCATTGATGATA 900
 QY 901 AAAGACTTCAACCTTCTGAAGTGAATCAACGGCAATTTCTTACAGACCTCTCACTATCCT 960
 Db 901 AAAGACTTCAACCTTCTGAAGTGAATCAACGGCAATTTCTTACAGACCTCTCACTATCCT 960
 QY 961 TACAGTGAAGAGTGGCTGATCTTGGCGACAGACTCGGAATCCTTGTGATAGAGAAACC 1020
 Db 961 TACAGTGAAGAGTGGCTGATCTTGGCGACAGACTCGGAATCCTTGTGATAGAGAAACC 1020
 QY 1021 CCGCAGCTTGGTATCAACAGGTACCACTACAAATCCCGAGACTCAGAAATAGCAGAAAGAC 1080
 Db 1021 CCGCAGCTTGGTATCAACAGGTACCACTACAAATCCCGAGACTCAGAAATAGCAGAAAGAC 1080
 QY 1081 AACATTAAGAGAAATGATGCACAGACACAAAGAACATCCAGTGTGATCATGTGAGTGTG 1140
 Db 1081 AACATTAAGAGAAATGATGCACAGACACAAAGAACATCCAGTGTGATCATGTGAGTGTG 1140
 QY 1141 GCGAAGCAACCAAGAGTCCAAACCATCCAGACGGGAGGTTCTTCAAAAGCCCTTTATGAG 1200
 Db 1141 GCGAAGCAACCAAGAGTCCAAACCATCCAGACGGGAGGTTCTTCAAAAGCCCTTTATGAG 1200

```
QY 1201 ACTGCCAATGAAATGGATCGAAACAGCCCCGGTTGTCATGTGTGAGCATGTAGACGACCA 1260
    |||||||
Db 1201 ACTGCCAATGAAATGGATCGAAACAGCCCCGGTTGTCATGTGTGAGCATGTAGACGACCA 1260
QY 1261 GACGAGAAACAAGACGCTGGCGCTGAAGTACTTCGACATCGTGTGTGAACAGGTAC 1320
    |||||||
Db 1261 GACGAGAAACAAGACGCTGGCGCTGAAGTACTTCGACATCGTGTGTGAACAGGTAC 1320
QY 1321 TACGGCTGTACATCTATCAGGGAAGATAGAAGAGACTTCAAGCTCTGAAAAAGAC 1380
    |||||||
Db 1321 TACGGCTGTACATCTATCAGGGAAGATAGAAGAGACTTCAAGCTCTGAAAAAGAC 1380
QY 1381 ATAGAAGAGCTCTATGCAAGGACAGAAAAGCCCATCTTTGTCAAGAATTCCGTGCGGAC 1440
    |||||||
Db 1381 ATAGAAGAGCTCTATGCAAGGACAGAAAAGCCCATCTTTGTCAAGAATTCCGTGCGGAC 1440
QY 1441 GCGATAGCTGCATCCACTACGATCCACTCAAAATGTTCTCCGAAGAGTACCAAGCAGAG 1500
    |||||||
Db 1441 GCGATAGCTGCATCCACTACGATCCACTCAAAATGTTCTCCGAAGAGTACCAAGCAGAG 1500
QY 1501 CTGCTTGA AAAAGAGATCAGGCTCCTTTTGA AAAAAGACTACATCATCGGAACAGCTG 1560
    |||||||
Db 1501 CTGCTTGA AAAAGAGATCAGGCTCCTTTTGA AAAAAGACTACATCATCGGAACAGCTG 1560
QY 1561 TGGGCTTTTGAGATTTTAAAGCTCCTCAGAATGTGAAGAACCCATTTCTAACACAGAAG 1620
    |||||||
Db 1561 TGGGCTTTTGAGATTTTAAAGCTCCTCAGAATGTGAAGAACCCATTTCTAACACAGAAG 1620
QY 1621 GGTTGTTTCAACAAGACAGACAAACCCAACTCGTTGCTCATGTACTGAGAAGACTGTGG 1680
    |||||||
Db 1621 GGTTGTTTCAACAAGACAGACAAACCCAACTCGTTGCTCATGTACTGAGAAGACTGTGG 1680
QY 1681 AGTGAGGTT 1689
    |||||||
Db 1681 AGTGAGGTT 1689

RESULT 3
ACN43424
ID ACN43424 standard; cDNA; 2050 BP.
XX
AC ACN43424;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic polynucleotide SEQ ID NO:2299.
XX
KW se; gene; gene therapy; human diagnostic and therapeutic polynucleotide;
KM dithp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
XX
PR 12-SEP-2002; 2002US-0410260P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstein EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
```

```
DR WPI; 2004-329368/30.
DR P-PSDB; ABM84772.
XX
PT New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
PS Claim 1, Page; 190pp; English.
XX
CC The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp polynucleotide of
CC the invention. Note: The sequence data for this patent is not represented
CC in the printed specification, but was obtained in electronic format
CC directly from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
SQ Sequence 2050 BP; 465 A; 532 C; 598 G; 455 T; 0 U; 0 Other;

Query Match 10.3%; Score 174; DB 13; Length 2050;
Best Local Similarity 52.2%; Pred. No. 2.7e-43;
Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;

QY 737 TGAACCTTGA AAAAAGACGAGTACACTCTGACATCGGAATCAGAAGATCAGCTGGAGC 796
    |||||
Db 815 TGGGGCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGACTGTGGCTGTACCA 874
QY 797 AGAAGAGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGAAACACAGAG 856
    |||||
Db 875 AGAGCCAGTTCTCATCAATGGGAAACCTTCTATTCCAGGTTGTCAACAAGCATGAGG 934
QY 857 AATCCCCGTTCTGGGGCAGGGCACCTTTTATCCATTGATGATAAAGACTCAACCTTC 916
    |||||
Db 935 ATGCGGACATCCGAGGGAAGGCTTCGACTGGCCGCTGTGTGAAGACTTCAACCTGC 994
QY 917 TGAAGTGAATCAACGGGAATCTTTCAGGACCTCTCACTATCTTACAGTGAAGAGTGC 976
    |||||
Db 995 TTCGCTGCGTTGTGCCAACGCTTTCGTACCAAGCCACTACCCCTATGCAGAGAAAGTGA 1054
QY 977 TGGATCTTGCCGACAGACTCGGAATCTTGTGATAGACGAAGCCCCGACGTTGTATCA 1036
    |||||
Db 1055 TGCAAGTGTGTGACCCGCTATGGGATGTGTGATCATGATGATGTCCCGGCGTGGCTGG 1114
QY 1037 CAAGGTACCA---CTACAATCCGAGACTCAGAGATAGCAGAAGACAACATAAGAGAA 1093
    |||||
Db 1115 CGCTGCCGCA GTTCTTCAACAAGTTCCTGTCATCACCAATGCAGGTGATGAAAGAG 1174
QY 1094 TGATGCACAGACACAAGAACCATCCAGTGTGATCATGTGAGCTGTGGCGAAGCAACAG 1153
    |||||
Db 1175 TGGTGCCTAGGGAACAAGAACCAACCCCGCGTGTGATGTGTGTGGCCACGAGCCTG 1234
QY 1154 AGTCCCAACCATCCAGACGCGGAGGTTCTTCAAAAGCCCTTATGAGACTGCCAATGAAA 1213
    |||||
Db 1235 CGTCCACCTTGAATCTGCTGCTACTACTTGAAGATGATGCTCCTCACACCAATCCT 1294
QY 1214 TGGATGGAACACGCCCCGTTGTGATGTGAGCATGATGAGCAGCAGACAGAGAACAA 1273
    |||||
Db 1295 TGAACCCCTCCCGGCTGTGACCTTGTGAGCAACTTAACCTATGACG----- 1342
QY 1274 GAGACGTGGCGCTGAAGTACTTTCGACATCGTCTGTGTGAACAGGTACTACGGCTGTACA 1333
    |||||
Db 1343 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTGAACAGCTACTACTTGTGTATC 1402
```


Db 1615 ACCATCTGGGTCTGGATCAAAAAACGAGAAAATACGTGGTGGAGAGCTCATTTGGAATT 1674
QY 1568 TTGCAGATTTTAAGACTCCTCAGAAATGTGAGAAAGCCCATTTCTCAACCAAGGGTGT 1627
Db 1675 TTGCCGATTTTCATGACTGAACAGTCAACGACGAGAGTGTGGGAATAAAAAGGGATCT 1734
QY 1628 TCACAAGAGACAGACAAACCCAACTGCTGCTCATGTACTGAGA 1671
Db 1735 TCACTCGGACAGACAAACCAAAAGTGCAGCGTTCCTTTTGGCA 1778

RESULT 5
ABL04147
ID ABL04147 standard; cDNA; 2128 BP.

XX ABL04147;
AC ABL04147;
DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 6923.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX Drosophila melanogaster.
OS WO200171042-A2.
PN 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
PF 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
PR 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
PA Venter JC, Adams M, Li PWD, Myers EW;
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX P-PSDB; ABB60044.
DR New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 6923; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 2128 BP; 546 A; 511 C; 540 G; 531 T; 0 U; 0 Other;

Query Match 10.3%; Score 174; DB 4; Length 2128;
Best Local Similarity 51.6%; Pred. No. 2.8e-43;
Matches 492; Conservative 0; Mismatches 435; Indels 27; Gaps 3;

QY 734 AGGTGGAACCTTGAAAAAGACGAGTACACTTGACATCGGAATCAGAACGATCAGCTGGG 793
Db 1104 AGTTGGAATCTCTGACGAGATACCTACCGCCTCCAGTGGGATACGTAAGTTAGTTGGG 1163
QY 794 ACGAGAAAGAGGCTCTATCTGAACGGGAAACCTGCTTTTGAAGGCTTTGAAAGCAGC 853
Db 1164 ACAACGACAGCCTGCTGCTGAATGAAAAACCCCTTATCTGCGGGATTGGACGGCAGC 1223

QY 854 AGGAATTCCTCCCTTCTGAGGAGGACACCTTTTATCCATTGATGATAAAGACTTCAACC 913
Db 1224 AGGACTCCGATATCCGCGGGAAGGATGGAATGATGCGCTTCTGCTAGAGATTTTAACC 1283
QY 914 TTCTGAAGTGATCAACCGGAATCTTTACAGACCTCTCACTATCTTACAGTGAAGAGT 973
Db 1284 TGCTGAAGTGAGCTGAGCCATATGATATGCACTCTCACTATCTTATTCGGAAGAGT 1343
QY 974 GCGTGATCTTGCCGACAGACTCGGAATCTTTGTATAGCAAGCCCCGCAAGTTGGTA 1033
Db 1344 CAATGCAGTTTGCCGATCAGCATGGTATTAATGATTATTGACGAATGCC-----TGCTG 1397
QY 1034 TCACAAGGTACCACTACAATCCCGAGACTCAGAAAGATAGCAAGAACATAAGAAAGAA 1093
Db 1398 TCAATATAGATATCTTGCAGCCGACGCTACTGAGAATCAGATGTCTCGTGAACAAC 1457
QY 1094 TGATCGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGCGAACGAACAG 1153
Db 1458 TGATCCACCGGACAGAAACCATCCAAGTGTGTTGATGTCGTTAGCCAACGAGCCGA 1517
QY 1154 AGTCCAACCATCCAGACGCGGAGGTTTCTTCAAAAGCCCTTATGAGACTGCCAATGAAA 1213
Db 1518 GATCGAACACAGGAGGAGCCCTTAATACCTTGAATTCCTGGTAAACTATGTAAGAGAAA 1577
QY 1214 TGATCGAACACGCCCCGTTGTCATGTGTGATGATGATGAGCCACAGACGAGAACAA 1273
Db 1578 TAGCTCACGACGACCTCTAACCGCGCTATAATGCCAATC-----TTCCA 1625
QY 1274 GAGACGTGGCGTGAAGTACTTCGACATCGTCTGTGTGAACGGTACTACGGCTGTGACA 1333
Db 1626 GTTGCCATTTGGCGCAGTTTCTGAGACATCGTGGGTTCAATCGCTATAACTCTTGGTATC 1685
QY 1334 TCTATCAGGAAGGATAGAGAAGACTTCAAGCTCTGAAAAAAGACATAGAAGAGCTCT 1393
Db 1686 AGAAGCTCGGCGCTACTGACATGATACTTAACCTGTTGACGATAGAGGCCAGAGTTGGC 1745
QY 1394 ATGCAAGGACAGAAAGCCCATCTTTGTCAAGAAATTCGGTCCGAGCGGATGACTGGCA 1453
Db 1746 GGGATAGTTCGGAAGCGCTGTCATCCAAATTTGAAGTACGGCGCGACACTATGAGGGCA 1805
QY 1454 TCCACTACGATCCACCTCAATGTCTCCGAAGAGTACCAGCAGAGCTCGT----- 1505
Db 1806 TGCACTCACTCCCGCCTTATTTGTGTCGAGGAATACAGGTTGAGCTCTTCTCCGCC 1865
QY 1506 -TGAAAAGACGATCAGGCTCTTTTGAAAAAAGACTACATCATCGGAACACACGTGTGG 1564
Db 1866 ATTTCAAGGCTTTCGACGAGTTACGCGGAGAGAGATGTTATTTGAGAGTTTGTGGA 1925
QY 1565 CTTTGCAGATTTTAAGACTCTCAGAAATGTGAGAAGACCCATTCTCAACCAAGGCTG 1624
Db 1926 ACTTCGCCGATTTTCGAGCCGCGCAGACTTAACCCGCGTGGCGGCAAAAAAGGAG 1985
QY 1625 TTTTCACAAGACAGACACACCCCAACTCGTTGCTCATGTACTGAGAAGACTGT 1678
Db 1986 TCTTTACAAGAACCGACACACCCCAAGAGTGGCTCAATTTTAAAGCGCGCGT 2039

RESULT 6
AAD50922
ID AAD50922 standard; DNA; 2169 BP.

XX AAD50922;
DT 02-APR-2003 (first entry)

XX l. mexicana SAP-human beta-GUS-IGF-II chimeric DNA.

XX lysosome; metabolic disease; lysosomal storage disease; gene therapy;
KW Gaucher's disease; Pompe's disease; Hurler's syndrome; neuroprotective;
KW Niemann-Pick's disease; Schindler's disease; mucopolidiosis; cystinosis;
KW Batten's disease; prosaposin; infantile neuronal ceroid lipofiscinosis;
KW fucosidosis; mannosidosis; antilipemic; insulin-like growth factor 2;

Query Match	10.3%	Score 174;	DB 8;	Length 2169;
Best Local Similarity	52.2%	Pred. No. 2.8e-43;		
Matches 493;	Conservative 0;	Mismatches 430;	Indels 21;	Gaps 4;

737 TGGAACTTGAAGAGCAGTAGACTCTTGACATCGGATCAGACGATCAGCTGGAGC 796

Db	941	TGGGGCTGTGTCGTGACTTCTAACACACTCCCTGTGGGATCCGCACCTGTGGCTGTACCA	1000
Qy	797	AGAAAGAGCTCTATCTGAAACGGGAAACCTGTCTTTTGAAGGGCTTTGGAAACAGCAGG	856
Db	1001	AGAGCCAGTTCTTCATCAATGGGAAACCTTTCTATTTCACGGTGTCAACAGCATGAGG	1060
Qy	857	AATTCCCCGTTCTGGGGCAGGGCACCTTTTATCCATTGTGATGAATAAAAGCTTCAACCTTC	916
Db	1061	ATGCGGACATCCGAGGAAAGGGCTTCGACTGTGGCCGCTGTGGTGAAGGACTTCAACCTGC	1120
Qy	917	TGAAGTGAATCAACGCGAATCTTTCAAGAACCTCTCACTATCTTACAGTGAAGTGGC	976
Db	1121	TTGCTGGCTTGTGTGCAAGCGCTTTCGTAACAGCCACTACCCCTATGACAGAAAGTGA	1180
Qy	977	TGGATCTTGCCGACAGACTGCGAATCTTGTGATAGACGAAGCCCCGACGTTGTATCA	1036
Db	1181	TGCAGATGTGTACCGCTATGGGATTGTGTTCATGATGAGTGTCCCGCGTGGTCTGG	1240
Qy	1037	CAAGGTACCA--CTACAAATCCCGAGACTCAGAAGATAGCAAGAACAACATAAGAGAA	1093
Db	1241	CGCTGCCGAGTTCTTCAACAGCTTCTCTGCATCACCAATGACAGTGAATGAAGAAG	1300
Qy	1094	TGATCGACAGACACAAGAACCATCCCAAGTGTATCATGTGAGAGTGTGGCGAACGAACCA	1153
Db	1301	TGGTGCATGGGACACAAGAACCAACCCCGGCTGTGATGTGTCTGTGGCCAAAGAGCTTG	1360
Qy	1154	AGTCCAAACCATCCAGACGGCGAGGGTTCTTCAAGCCCTTATGAGACTGCCAATGAAA	1213
Db	1361	CGTCCACCATAGAATCTGTGGCTACTACTTGAAGATGGTGTATGCTCACACCAATCTCT	1420
Qy	1214	TGGATCGAACACAGCCCCGTTGTCTATGTGTGAGCATGTGAGCGCACAGACGAGAAACAA	1273
Db	1421	TGGACCCCTCCCGGCTGTGACCTTTGTGAGCAACTCTAATATGACAG-----	1468
Qy	1274	GAGACGTGGCGCTGAAGTACTTTCGACATCGTCTGTGTGAACAGTACTACGGCTGTACA	1333
Db	1469	CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGTACTACTCTTGTATATC	1528
Qy	1334	TCTATCAGGGAAGATAGAGAAGACTTCAAGCTCTGGAAGAAAGACATAGAAGAGCTCT	1393
Db	1529	ACGACTACGGGACACCTGGAGTTGATTCACTGACGTGAGCCACCCAGTTTGAAGACTGGT	1588
Qy	1394	ATGCAAGGCACAGAAAGCCCATCTTTGTGACAGAAATTCGGTGCAGACGCGATAGCTGGCA	1453
Db	1589	ATTAAGAAATATCAGAAAGCCATTATTACAGAGCGATATGAGCAGAAACGATTGACAGGT	1648
Qy	1454	TCCACTACGATCCACCTCAAAATGTTCTCCGAAAGATACCAAGCAGAGCTCGTTGAAAAG-	1512
Db	1649	TTCACGAGATCCACCTCTGATGTTCACTGAAGATACCAAGAAAGTCTGCTAGACAGT	1708
Qy	1513	ACGATCAGGCTCCTTTTGAAGAA-----AGACTACATCATCGGAACACAGCTGTGGGCT	1567
Db	1709	ACCATCTGGGTCTGATCAAAAAACGCAAAATATGTGGTTGAGAGCTCATTTGCAATT	1768
Qy	1568	TTGCAGATTTTAAGACTCTCTCAGAATGTGAGAAGACCAATTCTCAACACAAGAGGTGTTT	1627
Db	1769	TTGCCGATTTTCATGACTGAACAGTACCGACGAGAGTGTGGGAATATAAAAGGGGATCT	1828
Qy	1628	TCACAAGAGACAGACAACCCAAACTGTTGCTCATGTACTGAGA	1671
Db	1829	TCACTCGGACAGACAAACAAAAGTGACGCTTCTTTTGGCA	1872
RESULT 7			
ADF47502			
ID	ADF47502 standard; cDNA; 2169 BP.		
XX	ADF47502;		
XX	26-FEB-2004 (first entry)		
DT	26-FEB-2004 (first entry)		
XX	Human beta-glucuronidase and IGF-II fusion protein cDNA SEQ ID NO:5.		
XX			

KW underglycosylated targeted therapeutic; human; lysosome;
KW lysosomal targeting domain;
KW cation-independent mannose-6-phosphate receptor;
KW underglycosylated therapeutic fusion protein; nephrotropic;
KW enzyme replacement therapy; cell therapy; gene therapy;
KW lysosomal storage disease; metabolic disease; enzyme defect;
KW Pompe disease; Tay-Sachs disease; Sandhoff disease; Fabry disease;
KW Gaucher disease; Krabbe disease; Wolman disease; Hurler syndrome;
KW Hunter syndrome; Sly syndrome; Schindler disease;
KW infantile sialic acid storage disease; Batten disease;
KW infantile neuronal ceroid lipofuscinosis; Ehlers-Danlos syndrome type VI;
KW congenital glycosylation disorder; fusion protein; gene; ss.
XX Synthetic.
OS Homo sapiens.
XX MO2003102583-A1.
XX 11-DEC-2003.
XX 29-MAY-2003; 2003WO-US017211.
XX 29-MAY-2002; 2002US-0384452P.
PR 05-JUN-2002; 2002US-0386019P.
PR 06-SEP-2002; 2002US-0408816P.
PR 16-OCT-2002; 2002US-00272531.
PR 06-FEB-2003; 2003US-0445734P.
XX (SYMB-) SYMBIONTICS INC.
XX Lebowitz JH, Beverley SM, Sly WS;
PI WPI; 2004-035473/03.
DR P-PSDB; ADFA7503.
XX Novel underglycosylated targeted therapeutic comprising therapeutic agent
PT active in human lysosome, lysosomal targeting domain binds to human
PT cation-independent mannose-6-phosphate receptor, for treating Pompe
PT disease.
XX Example 1; SEQ ID NO 5; 137bp; English.
PS The present invention describes an underglycosylated targeted therapeutic
XX (GT) which comprises: (a) a therapeutic agent that is therapeutically
CC active in human lysosome; and (b) a lysosomal targeting domain that binds
CC an extracellular domain of human cation-independent mannose-6-phosphate
CC receptor (CM) and (i) does not bind a mutin in which amino acid 1572 of
CC CM is changed from isoleucine to threonine; and (ii) binds the mutin
CC with dissociation constant at least ten times the dissociation constant
CC for binding CM. Also described: (1) an underglycosylated therapeutic
CC fusion protein (FP) comprising a therapeutic domain and a subcellular
CC targeting domain that binds to an extracellular domain of a receptor on
CC an exterior surface of a cell, and upon internalisation of the receptor,
CC permits localisation of the therapeutic domain to a subcellular
CC compartment where the therapeutic domain is therapeutically active; and
CC (2) production of GT. GT has nephrotropic activity, and can be used in
CC enzyme replacement therapy, cell therapy and gene therapy. The FP can be
CC used for treating a lysosomal storage disease patient by administering FP
CC to the patient. GT can be used for treating a patient by identifying a
CC targeting moiety that binds CM in a mannose-6-phosphate independent
CC manner, synthesising GT comprising a therapeutic agent that is
CC therapeutically active in a mammalian lysosome and a targeting moiety
CC that binds CM in a mammalian lysosome and a targeting moiety
CC administering GT to the patient, where the targeting moiety is identified
CC by screening a nucleic acid or peptide library. GT is useful for treating
CC metabolic disease, lysosomal storage diseases and associated enzyme
CC defects such as Pompe disease, Tay-Sachs disease, Sandhoff disease, Fabry
CC disease, Gaucher disease, Krabbe disease, Wolman disease, Hurler
CC syndrome, Hunter syndrome, Sly syndrome, Schindler disease, infantile
CC sialic acid storage disease, Batten disease, infantile neuronal ceroid
CC lipofuscinosis, Ehlers-Danlos syndrome type VI and congenital disorders
CC of glycosylation. The present sequence is used in the exemplification of
CC the present invention.

XX SQ Sequence 2169 BP; 468 A; 609 C; 637 G; 455 T; 0 U; 0 Other;
Query Match 10.3%; Score 174; DB 12; Length 2169;
Best Local Similarity 52.2%; Pred. No. 2.8e-43;
Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;
QY 737 TGGAACTTGA AAAAGACGAGTACACTCTGACATCGGAATCAGAACGATCAGCTGGAGC 796
DB 941 TGGGGCTGTGTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGGCTGTACCA 1000
QY 797 AGAAGAGCTCTATCTGAACGGGAAACCTGCTTTTGAAGGCTTTGAAAGCAGAG 856
DB 1001 AGAGCCAGTTCCTCATCAATGGGAAAACCTTCTATTCCACGGTGTCAACAAGCATGAG 1060
QY 857 AATTCCCGTTCGGGCGAGGCACTTTTATTCATTGATGATTAAGAAGCTTCAACTTC 916
DB 1061 ATCGGACATCCGAGGGAAGGCTTCGACTGGCCGCTGCTGTGAAGACTTCAACTGC 1120
QY 917 TGAAGTGAATCAACGGAATCTTTACAGACCTCTCACTATCTTACAGTGAAGTGC 976
DB 1121 TTCGCTGGCTGTGGTCCAAAGCTTCCGTACCAAGCACTACCTTATGACAGGAAGTGA 1180
QY 977 TGGATCTTCCGACAGACTCGGAATCTTGTGATAGACGAAGCCCGCAGCTGTATCA 1036
DB 1181 TGCAGATGTGACCCGCTATAGGATGTGTGATCATGATGATGTCCGCGGTGGTCTGG 1240
QY 1037 CAAGTACCA---CTACAATCCGAGACTGAGAAGATAGCAAAAGACAATAAGAAGA 1093
DB 1241 CGCTGCCGAGTTCTTCAACAACGTTTCTGCATCACCAATGACAGTGAAGAAGAAG 1300
QY 1094 TGATCGACAGACAAGAACCATCCAGTGTATCATGTGAGTGTGCGAAGCAACAG 1153
DB 1301 TGGTGCATAGGACAAGAACCACCCGCGGTGTGATGTGTGCGCAAGCAAGCTG 1360
QY 1154 AGTCCAACCATCCAGACGCGGAGGTTCTTCAAAAGCCCTTATGAGACTGCCAATGAA 1213
DB 1361 CGTCCACCTAGAAATCTGTGCTACTACTTGAAGATGTGATCGCTCACACCAATCCT 1420
QY 1214 TGGATGGAACAGCCCGCTTGTATGTGAGATGATGAGCGAACGAGAGAACAA 1273
DB 1421 TGAACCCCTCCCGGCTGTGACCTTTGTAGCAACTTAATATGACG----- 1468
QY 1274 GAGACGTGGCGTGAAGTACTTCGACATCGTGTGTAAGAGTACTACGGCTGTGACA 1333
DB 1469 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTGAACGCTACTACTCTTGTATC 1528
QY 1334 TCTATCAGGGAAGATAGAGAAGACTTCAAGCTCTGMAAAAAGACATAGAAGACTCT 1393
DB 1529 ACGACTACGGGCACTGAGTGTATTCAGCTGACAGCTGGCCACCAAGTTGAGAACTGT 1588
QY 1394 ATGCAAGGCACAGAAGCCCATCTTGTCAAGAAATTCGGTGGGACGGATAGCTGGCA 1453
DB 1589 ATAAGAAGTATCAGAAGCCCATTTATTCAGAGGAGTATGAGCAGAAACGATTGACAGGT 1648
QY 1454 TCCACTACGATCCACTCAATGTCTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAG- 1512
DB 1649 TTCAACGAGATCCACTCTGTGTCTACTGTAAGAGTACCAAGAAAGTCTGCTAGACAGT 1708
QY 1513 ACGATCAGGCTCTTTGAAAAA-----AGACTACATCATCGGAACACACGTTGGGCT 1567
DB 1709 ACCATCTGGGCTCTGATCAAAAACGAGAAATATGTGTTGAGAGAGCTCATTTGGAATT 1768
QY 1568 TTGCAGATTTTAAGACTCCTGGAATGTGAGAAAGACCCATTCTCAACCAAGGTTT 1627
DB 1769 TTGCCGATTTCATGACTGAACAGTCAACGACAGAGTGTGCGGAATTAAGGGGATCT 1828
QY 1628 TCACAAGAGACAGAACCAACCACTGTTGCTCATGTACTGAGA 1671
DB 1829 TCACTCGGCGAGAGACAACCAAAAAGTGACAGCTTCTTTGGCA 1872

RESULT 8

ADJ58608
 ID ADJ58608 standard; DNA; 2169 BP.
 XX AC ADJ58608;
 XX DT 06-MAY-2004 (first entry)
 XX DE Fusion DNA used in the exemplification of the invention.
 XX KW Targetted therapeutic; lysosome storage disease; gene therapy; human;
 KW IGF-II; insulin-like growth factor 2; beta-glucoronidase; GUS; chimeric;
 XX gene; ds.
 OS Homo sapiens.
 OS Chimeric.
 OS Unidentified.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..2169
 FT /tag= b
 FT /product= "Fusion protein"
 FT 1..69
 FT /tag= a
 FT /note= "Signal peptide DNA"
 FT 70..1956
 FT /tag= c
 FT /note= "Human mature P-glucoronidase DNA"
 FT 1957..1965
 FT /tag= d
 FT /note= "Encodes an amino acid bridge"
 FT 1966..2166
 FT /tag= e
 FT /note= "Human IGF-II DNA"
 XX
 PN US2004006008-A1.
 PD 08-JAN-2004.
 XX
 PF 16-OCT-2002; 2002US-00272483.
 XX
 PR 30-APR-2001; 2001US-0287531P.
 PR 10-JUL-2001; 2001US-0304609P.
 PR 15-OCT-2001; 2001US-0329461P.
 PR 23-JAN-2002; 2002US-0351276P.
 PR 30-APR-2002; 2002US-00136841.
 PR 29-MAY-2002; 2002US-0384452P.
 PR 05-JUN-2002; 2002US-0386019P.
 PR 06-SEP-2002; 2002US-0408816P.
 XX
 PA (SYMB-) SYMBIONTICS INC.
 XX
 PI Lebowitz JH, Beverley SM;
 XX WPI; 2004-081736/08.
 DR P-PSDB; ADJ58609.
 XX
 XX
 PT New targeted therapeutic comprising a therapeutic agent that is active in
 PT mammalian lysosome, and means for binding a domain of human cation-
 PT independent mannose-6-phosphate receptor, useful for treating lysosome
 PT storage disease.
 XX
 PS Example 1; SEQ ID NO 5; 46bp; English.
 XX
 CC The invention relates to a targeted therapeutic which comprises a
 CC therapeutic agent that is active in a mammalian lysosome and means for
 CC binding an extracellular domain of human cation-independent mannose-6-
 CC phosphate receptor in a mannose-6 phosphate-independent manner. The
 CC targeted therapeutic, therapeutic fusion protein and methods of the
 CC invention are useful for treating lysosome storage diseases. The
 CC invention is also useful in gene therapy. The present sequence is a
 CC fusion DNA used in the exemplification of the invention. This DNA encodes
 CC a fusion protein which comprises a signal peptide sequence, a human
 CC mature beta-glucoronidase (GUS) sequence, a bridge of three amino acids

CC and a human IGF-II sequence.
 XX
 SQ Sequence 2169 BP; 468 A; 609 C; 637 G; 455 T; 0 U; 0 Other;
 Query Match 10.3%; Score 174; DB 12; Length 2169;
 Best Local Similarity 52.2%; Pred. No. 2.8e-43;
 Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;

QY 737 TGGAACTTGA AAAAGACGATGACCTGTGACATCGGAATCAGATCAGCTGGAGC 796
 DB 941 TGGGGCTGTGTGACTTCTTACACTCCCTGTGGGATCCGACTGTGCTGCACCA 1000
 QY 797 AGAAGAGCTCTATCTGAACGGAAACCTGTCTTTTGAAGGCTTTGMAAGCAGAG 856
 DB 1001 AGAGCCAGTTCCTCATCATGGAACCTTTCTATTTCACCGGTGTCAACAGCATGAG 1060
 QY 857 AATTCCCGTTCTGGGCGAGGCACTTTTATCCATTGATGATTAAGACTTCACTTC 916
 DB 1061 ATGCGGACATCCGAGGGAAGGCTTCGACTGGCCGCTGTGTAAGACTTCACTGC 1120
 QY 917 TGAAGTGATCAACGGAATTTCTCAGACCTCTCACTATCTTACGTGAAGATGGC 976
 DB 1121 TTGCGCTGTGTGTCACACGCTTCCGTACAGCCACTACCCCTATGACAGGAATGA 1180
 QY 977 TGGATCTTGCCGACAGACTCGGAATCCTTGATGATAGCAAGCCCGCAGCTGTATCA 1036
 DB 1181 TGCAGATGTGTACCGCTATGGGATGTGTGATCATGATGATGTCCGCGTGGTCTGG 1240
 QY 1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAGATAGCAGAAACAATAGAAGAA 1093
 DB 1241 CGCTGCCGAGTCTTCAACAACGTTTCTGTCATCACCACATGCAAGTGATGAAGAAG 1300
 QY 1094 TGATCGACAGACAAAGAACCAATCCAGTGTGATCATGTGAGTGTGGGAACAGCAG 1153
 DB 1301 TGGTGCCTAGGGAACAAGAACCCCGCGTGTGATGTGTGTGGCCAAAGCCTG 1360
 QY 1154 AGTCCAACCATCAGACGCGGAGGTTTCTTCAAGCCCTTATGAGACTGCCAATGAAA 1213
 DB 1361 CGTCCACCTAGAAATCTGTGCTACTACTTGAAGATGTGTATCGCTCACACCAATCCT 1420
 QY 1214 TGGATCGAACACGCCCCGTGTGATGTGAGCATGATGACCGACCAAGCAGAGAACA 1273
 DB 1421 TGAACCCCTCCCGGCTGTGACCTTTGTGAGCACTTAATATGACAG----- 1468
 QY 1274 GAGACGTGCGCTGAAGTACTTGCACATCGTCTGTGTGAACAGTACTACGCGTGTACA 1333
 DB 1469 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTGAACAGTACTACTCTTGGTATC 1528
 QY 1334 TCTATCAGGGAAGATAGAAGAGACTTCAAGCTCTGAAAAAGACATAGAAGCTCT 1393
 DB 1529 ACGACTACGGGCACTTGAAGTATTCAGCTGCACTGGCCACCAAGTTTGAAGACTGGT 1588
 QY 1394 ATGCAAGGACAGAAAGCCCATCTTGTACACGAATTGGGTGCGAAGCGATAGCTGGCA 1453
 DB 1589 ATAAAGAGTATCAGAAAGCCCATTAATTCAGAGCGAGATAGGACAGAAACGATTGCAAGGT 1648
 QY 1454 TCCACTACGATCCACTCAATGTCTCCGAAGATACCAAGCAGAGCTCGTTGAAAAG- 1512
 DB 1649 TTCAACAGGATCACTCTGATGTTCACGTAAAGATACCAAGAAAGTGTGCTAGACAGT 1708
 QY 1513 ACGATCAGGCTCTTTGAAAAA---AGACTACATCATCGGAACACACGTTGGGCT 1567
 DB 1709 ACCATCTGGGTCTGGATCAAAAACGAGAAAATATGTGTTGAGAGCTCATTTGGAATT 1768
 QY 1568 TTGCAGATTTTAAGACTCTCAGAAATGTGAAGAACCCATTCTCAACCAAGGGTGT 1627
 DB 1769 TTGCCGATTTTCACTGAAGTCAACGATCAACGAGAGAGTGTGGGGAATTAAGGGGATCT 1828
 QY 1628 TCACAAGAGACAGAACCAACCACTGTTGCTCATGTACTGAGA 1671
 DB 1829 TCACTCGGAGAGACAAACCAAAAGTGACAGCGTTCCTTTGGCA 1872

```

RESULT 9
AAD06386 standard; DNA; 1956 BP.
ID AAD06386 standard; DNA; 1956 BP.
XX
XX AAD06386;
AC
XX
XX 10-AUG-2001 (first entry)
DT
XX
XX Human beta-glucuronidase (GUS) DNA.
DE
XX
XX Human; adeno-associated viral expression vector; AAV; gene therapy;
KM lysosomal storage disease; LSD; mucopolysaccharidoses VII; MPS VII;
KM Sly syndrome; beta-glucuronidase; GUS; glycosaminoglycan; GAG; ds.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..1956
FT /*tag= a
FT /product= "Human beta-glucuronidase (GUS)"
XX
XX WO200136603-A2.
XX
XX 25-MAY-2001.
XX
XX 17-NOV-2000; 2000WO-US031688.
XX
XX 17-NOV-1999; 99US-0166097P.
XX 30-JUN-2000; 2000US-0215430P.
XX
XX (AVIG-) AVIGEN INC.
XX (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
XX
XX Podsakoff G, Watson G, Couto LB, Yang B;
XX
XX MPI: 2001-343814/36.
XX DR P-PSDB; AAE02443.
XX
XX Use of recombinant adeno-associated virus, comprising gene encoding a
PT protein defective or missing in lysosomal storage disease, in the
PT manufacture of a medicament for treating the lysosomal storage disease.
XX
XX Disclosure; Page 77-80; 97pp; English.
XX
XX The present invention relates to recombinant adeno-associated virus (AAV)
CC expression vectors and virions, which include genes coding for enzymes
CC defective or missing in lysosomal storage disease (LSD). AAV is useful in
CC the manufacture of a medicament for treating lysosomal storage disease
CC e.g., mucopolysaccharidoses VII (MPS VII). MPS VII (Sly Syndrome) is due
CC to the deficiency in the lysosomal enzyme beta-glucuronidase (GUS) which
CC aids in the breakdown of glycosaminoglycans (GAGs). AAV is used in gene
CC therapy. The present sequence is human GUS DNA. This sequence is used in
CC AAV constructs
XX
XX Sequence 1956 BP; 437 A; 533 C; 563 G; 423 T; 0 U; 0 Other;
SQ

```

```

Query Match 10.2%; Score 172.4; DB 4; Length 1956;
Best Local Similarity 52.1%; Pred. No. 8.4e-43;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

```

```

OY 737 TGAAGCTGAAAAAGACGAGTACCTCTGACATCGAATCAGACGATCAGCTGGAGCG 796
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 938 TGGGCGCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGCTGCACCA 997
OY 797 AGAAGAGGCTCTATCTGAACGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAGG 856
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 998 AGAGCCAGTTCCTCAATGGAACCTTTCTATTTCCACGGTGTCAACAAGCATGAGG 1057
OY 857 AATTCCTCGTTCTGGGCGAGGCGACCTTTTATCCATGATGATAAAGACTTCAACTTC 916
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1058 ATGCGGACATCCGAGGGAAGGCTTGACTGCGCTGTGTGAAGACTTCAACTTGC 1117
OY 917 TGAAGTGATCAACGCGAATTCTTTCAAGACCTCTCACTATCTTACAGTGAAGATGGC 976

```

```

DB 1118 TTCGCTGGCTTGTGCCAACGCTTTCGTTACCAACCACCTATATGACAGGAAGTGA 1177
OY 977 TGGATCTTGCCGACAGACTCGGAATCCTTGTGATAGACGAAGCCCCGACGTTGGTATCA 1036
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1178 TGCAATGTGTGACCGCTATGAGGATTTGTGTATCATCATGATGATGTCCGCGTGGCCCTGG 1237
OY 1037 CAAGGTACCA---CTACAATCCCGAGACTCGAAGATAGCAGAACAACATAGAAGAA 1093
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1238 CGCTGCCGAGTTCTTCAACAACGTTTCTGTGCATCACACATGCAAGTGTATGAAGAA 1297
OY 1094 TGATCGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAACGACAG 1153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1298 TGTGCGTAGGAGACAAGAACCAACCCGCGGTGTGATGTGTGTGCGCAACGACCTG 1357
OY 1154 AGTCCAACCATCCAGACGCGGAGGTTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAAA 1213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1358 GGTCCACCTAGAAATCTGTGGCTACTACTTGAAGATGTGATCGCTCACACCAATCCT 1417
OY 1214 TGGATCGAACACGCCCCGTTGTCTATGTGAGCATGATGACGCCACGACGAGAGAACA 1273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1418 TGAACCCCTCCCGGCTGTGACCTTTGTAGCAACTCTAATATATGACAG----- 1465
OY 1274 GAGACGTGGCGCTGAAGTACTTGCACATCGTCTGTGTGAACAGTACTACGGCTGTACA 1333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1466 CAGACAAAGGGGCTCCGATATGTGATGTGATCTGTTGAACGCTACTACTCTTGTATC 1525
OY 1334 TCTATCAGGGAAGATAGAGAAGACTTCAAGCTCTGAAAAAAGACATAGAAGACTCT 1393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1526 ACGACTACGGGCACTGTAGTTGATTCAAGCTGACAGCTGGCCACCGAGTTGAGAACTG 1585
OY 1394 ATGCAAGGACAGAAAGCCCATCTTTGTCAAGAATTCCGTCGGACGCGATAGCTGCA 1453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1586 ATAGAAGTATCAGAAGCCCATTTATTCAGAGGAGATATGAGACAGAAACGATTCAGAG 1645
OY 1454 TCCACTACGATCCACCTCAATGTCTCCGAGAAGTACCAAGACAGAGCTCGTTGAAAAG- 1512
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1646 TTCACCAAGATCCACCTCTGATGTTCATGAGAGTACCAAGAAAGTCTGCTAGACAGT 1705
OY 1513 ACGATCAGGCTCCTTTTGA AAAA-----AGACTACATCATCGGAACACAGTGGGCT 1567
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1706 ACCATCTGGGCTGTGATCAAAAACGAGAAATATGTGTTGAGAGAGCTCATTTGAATT 1765
OY 1568 TTGCAGATTTTAAGACTCCTCAGATGTGAGAAAGACCATTTCAACACAGGCTGTTT 1627
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1766 TTGCCGATTTCTATGACTGAACAGTCAACCGACAGAGAGTGTGGGAATAAAGGGGATCT 1825
OY 1628 TCACAAGAGACAGACAAACCAAACTCGTTGCTCATGTACTGAGA 1671
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1826 TCACTCGGCGAGAGACAACCAAAAAGTGACAGCTTCTTTTGCA 1869

```

```

RESULT 10
ABK83932
ID ABK83932 standard; cDNA; 2191 BP.
XX
XX ABK83932;
AC
XX
XX 14-AUG-2002 (first entry)
DT
XX
XX Human cDNA differentially expressed in granulocytic cells #503.
DE
XX
XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
KM viral infection; parasitic infection; protozoal infection;
KM fungal infection; sterile inflammatory disease; psoriasis;
KM rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KM cardiac reperfusion injury; renal reperfusion injury; ARDS;
KM adult respiratory distress syndrome; inflammatory bowel disease;
KM Crohn's disease; ulcerative colitis; periodontal disease;
KM granulocyte activation; chronic inflammation; allergy.
XX
XX Homo sapiens.
XX

```

PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US030821.
XX
PR 03-OCT-2000; 2000US-0237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX WPI; 2002-435328/46.
DR
XX
PT Detecting granulocyte activation by detecting differential expression of
PT genes associated with granulocyte activation, which serves as diagnostic
PT markers that is useful for monitoring disease states and drug toxicity.
XX
PS Claim 1; SEQ ID NO 503; 114pp; English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing the
CC expression level to an expression level in an unactivated GC, where
CC differential expression of Gs is indicative of GCA. Also included are
CC modulating (M2) GA by contacting GC with an agent that alters the
CC expression of at least one gene in Gs; (2) screening (M3) for an agent
CC capable of modulating GCA or an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease using the gene expression
CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease, by detecting the level of
CC expression in a sample of the tissue of gene(s) from Gs, where the level
CC of expression of the gene is indicative of inflammation; (4) treating
CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease, by contacting a tissue having inflammation with an
CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC for screening an agent capable of modulating GCA preferably in an
CC inflammation in a tissue; M4 is useful for detecting an inflammation
CC (especially chronic) in a tissue, an allergic response in a subject,
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC peoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease; also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above
CC conditions. The present sequence represents a gene differentially
CC expressed in granulocytes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2191 BP; 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;

Query Match 10.2%; Score 172.4; DB 6; Length 2191;
Best Local Similarity 52.1%; Pred. No. 9e-43;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

QY 737 TGGAACTTGAAAAAGCAGTACACTCTGACATCGGAATCAGAAGCATCAGTGGGACG 796
DB 964 TGGGGCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGACTGTGCTGTACCA 1023
QY 797 AGAAGAGGCTCTATCTGAACGGGAACCTGCTTTTGAAGGGCTTTGGAAGCAGCAGG 856
DB 1024 AGAGCAGTTCTCTCATCAATGGGAACCTTTCTATTTCACGGGTGTCAACAGCATGAGG 1083
QY 857 AATTCCCGCTTCTGGGCGAGGGCACCCTTTTATTCATGTATGATAAAGACTTCAACCTTC 916
DB 1084 ATGCGGACATCCGAGGGAAGGGCTTGCACTGCGCTGTGCTGTAAGGACTTCAACCTGC 1143

QY 917 TGAAGTGATCAACGCGAATTCTTTAGAGACCTTCATCTATCTTACAGTGAAGATGCC 976
DB 1144 TTCCGTGGCTTGGTCCCAACGCTTTCCGTACCAGCCACTACCCCTATGCAAGAGAAATGA 1203
QY 977 TGGATCTTGGCGACAGACTCGGAATCTTGTGATAGACGAAGCCCCGACAGTTGATCA 1036
DB 1204 TGCAGATGTGTGACCGCTATATGGATTTGTGTCATCGATGAGTGTCCCGGCGTGGCTGG 1263
QY 1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAGAATAGCAAGAACACATAAGAGAA 1093
DB 1264 CGCTGCCGCAAGTTCTTCAACAACGTTTCTCTGCATCACACATGCAGGTGATGAAGAA 1323
QY 1094 TGATCAGACAGACACAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGCAACAG 1153
DB 1324 TGGTGGCTAGGGACAGAACCACCCCGGCTGTGATGTGTGCTGTGGCCAAAGAGCTTG 1383
QY 1154 AGTCCAACCATCCAGACCGGAGGGTTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAAA 1213
DB 1384 CGTCCCACTAGAAATCTGTGCTACTACTTGAAGATGTGTATCGCTCACACCAAAATCCT 1443
QY 1214 TGGATCGAACACGCCCCGTGTGATGTGATGATGAGATGAGCGCAGACAGAGAAACAA 1273
DB 1444 TGGACCCCTCCCGGCTGTGACCTTTGTGAGCAACTTAATATATGAG----- 1491
QY 1274 GAGACGTGGCGCTGAAGTACTTTCGATCGTCTGTGTGAACAGACTTACCGCTGTACA 1333
DB 1492 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGACTACTCTTGTGATC 1551
QY 1334 TCTATCAGGGAAGATAGAGAAGGACTTCAAGCTCTGAAAAAGACATAGAAAGCTCT 1393
DB 1552 ACGACTACGGGACACTGAGTTGATTCAGCTGCACTGCGCACCAGTTTGAGAACTGGT 1611
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTTGTACACAGATTCGGTGGCGACGATAGTGCA 1453
DB 1612 ATTAAGAAATATCAGAAAGCCCATTAATTCAGAGCGAGTATGAGCAGAAACGATTGAGGGT 1671
QY 1454 TCCACTACGATCCACTCAATGTCTCCGAAGAGTACCAAGCAGAGCTGTTGAAAAG- 1512
DB 1672 TTCAACAGAGATCCACTCTGATGTTCATGMAAGTACAGAAAGTCTGTAGAGCAGT 1731
QY 1513 ACGATCAGGCTCTTTGAAAAA-----AGACTACATCATCGGAACACACGCTGCGGCT 1567
DB 1732 ACCATCTGGGTCTGATCAAAAAGCAGAAAATATGTGTTGAGAGCTCATTTGAATT 1791
QY 1568 TTGCAGATTTTAAGACTCTCTCAGAATGTGAGAAGACCATTTCTCAACCAAGGATGTT 1627
DB 1792 TTGCCGATTTTCACTGAAGTGAACGAGTACCGAGAGTGTGGGAAATAAAAAGGAGATCT 1851
QY 1628 TCAAGAAGACAGACAACCCAACTGTTGCTCATGTACTGAGA 1671
DB 1852 TCACTCGGACAGACAACCAAAAAGTGACGCTTCCTTTTGGCA 1895

RESULT 11
ADD29602
ID ADD29602 standard; mRNA; 2191 BP.

AC ADD29602;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human tumour suppressor mRNA SEQ ID NO:55.
XX
KW ss; human; tumour suppressor; cancer; cancer; cytostatic; gene therapy.
XX
OS Homo sapiens.
XX
PN WO2003058201-A2.
XX
PD 17-JUL-2003.
XX
PF 31-DEC-2002; 2002WO-US041825.

31-DEC-2001; 2001US-0345317P.
(OUAR-) QUARK BIOTECH INC.
(CLEV-) CLEVELAND CLINIC FOUND.
Feinstein E, Gudkov AV;
WPI; 2003-598393/56.
Diagnosing cancer comprises determining the polypeptide or polynucleotide levels e.g., hepatic lipase, in a sample from a subject, where a higher level compared to that in a subject free of cancer is indicative of cancer.
Disclosure; SEQ ID NO 55; 272pp; English.
The invention relates to a novel method for diagnosing a cancer in a subject. the method comprises determining, in a sample from the subject, the level of at least one polypeptide, where a higher level of the polypeptide compared to the level of the polypeptide in a subject free of cancer is indicative of cancer. The polypeptide is selected from any of the polypeptides encoded by the polynucleotides listed in the specification and polypeptides which are at least 70% homologous to the polypeptides. The method of the invention has cytostatic activity, and may have a use in gene therapy. The method is useful in identifying markers specific for one or several types of cancer, depending on the tissue origin, which may be used in numerous diagnostic and prognostic applications as well as cancer type-specific targets for therapeutic intervention. The compounds that modulate the activity of a tumour suppressor gene are useful in the treatment of cancer or as anti-cancer drugs. The present sequence represents a polynucleotide of the invention.

Query Match 10.2%; Score 172.4; DB 10; Length 2191;
 Best Local Similarity 52.1%; Pred. No. 9e-43;
 Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4

Oy	737	TGGAATTGTAAAAAGACGAGTACA	CTTGAGACATCGAATCAGAACGATCAGCTGGGACG	796
Dp	964	TGGGGCCGTGTCTGACTTTCTACACACTCC	CTGTGGGGATCCGACTGTGGCTGTCA	1022
Oy	797	AGAAGAGGCCTTAATCTGAACCGGAAAC	CTGTCTTTTGAAGGGCTTTGAAAAGCAGAG	856
Dp	1024	AGAGCCAGTTCTTCATCAATGGGAACTTT	CTAATTTCCACGGTGTCAACAMGATGAGG	1083
Oy	857	AATTCGCCGTTCTGGGGCAGGGCAC	CTTTATCATGATGATAAAGACTTCAACTTC	916
Dp	1084	ATGCGGACATCCGAGGGAAGGGCTTTC	GACTGGCCGCTGTGTGAAGACTTCAACTTC	1144
Oy	917	TGAAGTGATCAACCGCAATTTCTTCA	GACCCTTCATCTTACAGTAGAAGATGGC	976
Dp	1144	TTGCGTGGCTTGTGTCCAACGCTTTC	CGTACCAGCACTAACCTTAATGAGAGAGTGA	1203
Oy	977	TGGAATCTTGCCGACAGACTCGAATCT	TGTGATAGACGAAGCCCCGACGTTGTATCA	1033
Dp	1204	TGCAGATGTGTAGCCGCTATGGAGAT	TGTGTGATCATGATGTCCCGCGTGGCCCTGG	1263
Oy	1037	CAAAGTACCA---CTACAATCCCAGACT	CCAGAAGATAGCAAAGACAATAGAGAA	1093
Dp	1264	CGGTGCCGAGTTCTTCAACAACGTTTT	CTCTGCATCACACATGCAAGTATGAAGAA	1322
Oy	1094	TGATCGACAGACACAAGAACCATCCCA	GTGTGATCATGTGGAGTGTGGCAACGAAC	1153
Dp	1324	TGTGTGCTAGGACAAGAACCAACCC	CGCGGTGTGTGTCTGTGGCCAAACGACCTG	1383
Oy	1154	AGTCCAAACCATCCAGACGCGAGGGTT	CTTCAAAAGCCTTATGAGACTGCCAATGAAA	1211
Dp	1384	CGTCCCACTAGAAATCTGTCTGCTACT	ACTTGAAGATGTGATCGCTCACACCAATCTCT	1443
Oy	1214	TGGATCGAACACGCCCCGTTGTCA	TGTGATGATGATGAGCGCACGACGAGAGAACAA	1277

CC specification also discloses an array comprising polynucleotides
CC hybridizing to the following genes: FOXM1, PRAME, Bcl2, STK15, CEGP1, K1-
CC 67, GSTM1, CA9, PR, BBC3, NME1, SURV, GAT3, TERC, YB-1, DYPD, GSTM3,
CC RPS6KB1, Src, Chk1, ID1, Esr1, p27, CCNB1, XIAP, Chk2, CDC25B, IGFBP3,
CC AKO55699, PI3KCA, TGFBR3, BAG1, CYP3A4, EPCAM, VEGFC, PS2, hENT1, WISP1,
CC HNF3A, NFkBp65, BRCA2, EGFR, TK1, VDR, Contig51037, pENT1, EPHX1, IFIA,
CC CDH1, HIF1c, IGFBP3, CTSEB, Her2 and DIABLO, immobilized on a solid
CC surface. The methods are useful for predicting clinical outcome for a
CC patient diagnosed with cancer, classifying cancer, and predicting the
CC likelihood of long-term survival of a breast cancer patient, or a patient
CC diagnosed with invasive breast cancer or with estrogen receptor (ER)-
CC positive invasive breast cancer. This sequence corresponds to a gene
CC sequence whose expression is detected by the method of the invention.
XX
SQ Sequence 2191 BP; 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;

Query Match 10.2%; Score 172.4; DB 10; Length 2191;
Best Local Similarity 52.1%; Pred. No. 9e-43;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

QY 737 TGGAACTTGAAGGAGAGTACCTTGACATCGGAATCAGAACGATCAGCTGGGACG 796
DB 964 TGGGGCTGTGTGCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGCTGCACCA 1023
QY 797 AGAAGAGCTCTATCTGAACGGGAACTGTCTTTTGAAGGGCTTTGGAAGCAGAGG 856
DB 1024 AGAGCCAGTTCCTCATCATGGAACCTTTCTATTCCACGGTGTCAACAAGCATGAGG 1083
QY 857 AATTCCCGCTTCTGGGGCAGGGACCTTTTATCCATGTGATATAAAGCTTCAACCTTC 916
DB 1084 ATGGCGACATCCGAGGGGAGGCTTGAAGTGGCGCTGTGTGAAGACTTCAACCTGC 1143
QY 917 TGAATGGATCAACGCGAATTTCTTCAAGACCTTCACTATCCTTACAGTGAAGATGGC 976
DB 1144 TTGCTGTGCTTGTGTGCAACGCTTCCGTACCAAGCACTACCCCTATGACAGAGAGTGA 1203
QY 977 TGGATCTTCCGACGAGACTCGAATCTTGTGATAGCAGAGCCCGCAGCTTGGTATCA 1036
DB 1204 TGCAGATGTGTGACCGCTATGGATGTGTGATGATGATGTGTCCCGGCTGG 1263
QY 1037 CAAGGTACCA---CTACAATCCGAGACTCAGAAGTAGCAGAGCAACATAAGAGAA 1093
DB 1264 CGTGCCCGCAGATTCTTCAACAACGTTTCTGTGATCAACACATGACGTGATGAAGAAG 1323
QY 1094 TGATCGACAGACACAAGAACCATCCCAAGTGTGATCATGTGAGTGTGGGAGCAACCAAG 1153
DB 1324 TGTGTCGTAGGAGCACAAGAACCAACCCCGCGGTCTGTGATGTGTGTGCCCAACGACCTG 1383
QY 1154 AGTCCAACCATCCAGACCGCGAGGGTCTTCAAAAGCCCTTATGAGACTGCCAATGAAA 1213
DB 1384 CGTCCCACTAGATCTGCTGCGTACTACTTGAAGATGTGATCGCTCACACCAAAATCCT 1443
QY 1214 TGGATCGAACACGCCCCGTTGTCTATGTGTGAGCATGATGAGCGCACGACGAGAGAACAA 1273
DB 1444 TGGACCCCTCCCGGCTGTGACCTTTGTGAGCAACTCTAATATGAG----- 1491
QY 1274 GAGACGTGGCGCTGAAGTACTTTCGACATCGTGTGTGAACAGGTAATAAGGCTGTACA 1333
DB 1492 CAGACAAAGGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGCTACTACTCTTGGTATC 1551
QY 1334 TCTATCAGGGAAGATAGAAGAGACTTCAAGCTCTGAAAAAGACATAGAAGAGCTCT 1393
DB 1552 ACGACTACGGGACACTGGAGTTGATTCAGCTGACGTGGCCACCCAGTTTGAAGAACTGGT 1611
QY 1394 ATGCAAGGCACAGAAAAGCCCATCTTTGTACAGAAATTCGGTGGGACGGATAGCTGGCA 1453
DB 1612 ATAGAGAGTATCAGAAAGCCCATTTATTCAGAGCGAGTATGAGCAGAAAACGATTGCAAGGT 1671
QY 1454 TCCACTACGATCCACCTCAATGTCTCCGAAAGATACCAAGCAGAGCTCGTTGAAAG- 1512
DB 1672 TTACCAAGGATCCACTCTGATGTTCACAGAGATACCAAGAAAGTCTGCTAGACAGT 1731
QY 1513 ACGATCAGGCTCCTTTTGAAGAA-----AGACTACATCATCGGAACACAGTGTGGGCT 1567

DB 1732 ACCATCTGGGTCTGATCAAAAACGAGAAATATGTGTTGGAGAGCTCATTTGGAATT 1791
QY 1568 TTGCAGATTTTAAGACTCCTCAGATGTGAGAGAACCCATTCTCAACCAAGAGGTGTTT 1627
DB 1792 TTGCCGATTTCATGACTGAACAGTCACCCGACGAGAGTGTGGGGAATAAAAGGGGATCT 1851
QY 1628 TCACAAGAGACAGACACACCCAACTCGTGTCTATGTACTGAGA 1671
DB 1852 TCATCGGACAGACACCAAAAGTGCAGCGTCTTTTGCGA 1895

RESULT 13
ACF57506
ID ACF57506 standard; DNA; 2191 BP.

XX ACF57506;
XX
XX 22-APR-2004 (first entry)

XX Human beta-glucuronidase DNA SEQ ID NO: 7.

XX Protein targeting; angiogenesis inhibitor; fusion gene; cancer; gene; ds;
KW antiangiogenic; cytostatic; ophthalmological; antiarthritic;
KW antirheumatic; antiinflammatory; dermatological; immunosuppressive;
KW antipsoriatic; antidiabetic; neuroprotective; vasotropic; anorectic.

OS Homo sapiens.
XX
PN WO2003093303-A1.

XX 13-NOV-2003.

XX 06-MAY-2003; 2003WO-US014243.

XX 06-MAY-2002; 2002US-0380063P.

XX (TEXA) UNIV TEXAS SYSTEM.

XX Hung M, Ian K, Ou-Yang F, Liu J, Ian K;

XX WPI; 2004-053038/05.

PT New polypeptide comprising an antiangiogenesis polypeptide region linked
PT to a therapeutic protein or a diagnostic protein, useful for diagnosing
PT or treating angiogenesis-dependent diseases e.g. cancer, lupus or
PT restenosis.

PS Disclosure; page 86; 0pp; English.

XX The present invention relates to a polypeptide comprising an
CC antiangiogenesis polypeptide region linked to a therapeutic protein or
CC polypeptide region, or a diagnostic protein or polypeptide region. The
CC polypeptide comprising an antiangiogenesis polypeptide region or the
CC nucleic acid encoding the polypeptide comprising the antiangiogenesis
CC polypeptide is useful for diagnosing or treating angiogenesis-dependent
CC diseases or cancer. Angiogenesis-dependent diseases include age-related
CC macular degeneration, atherosclerosis, angiofibroma, neovascular
CC glaucoma, arteriovenous malformations, nonunion fractures, arthritis,
CC rheumatoid arthritis, lupus, connective tissue disorders, Osler-Weber
CC syndrome, psoriasis, corneal graft neovascularisation, pyogenic
CC granuloma, delayed wound healing, retrolental fibroplasia, diabetic
CC retinopathy, scleroderma, granulations, haemangioma, trachoma,
CC haemophilic joints, vascular adhesions, hypertrophic scars, multiple
CC sclerosis, restenosis or obesity. The present sequence is a coding
CC sequence shown in the exemplification of the invention
XX

SQ Sequence 2191 BP; 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;

Query Match 10.2%; Score 172.4; DB 12; Length 2191;
Best Local Similarity 52.1%; Pred. No. 9e-43;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

QY	737	TGGA	ACTTGAAAAAGACGAGTACACTCTTGACATCGGAATCAGAACGATCAGCTGGGACG	796
Db	964	TGGGGC	CTGTGTCTGACTTCTTACACACTCCCTGTGGGGATCCGCAC	TGTGGCTGTCA
QY	797	AGAA	GAGGCTCTATCTGACCGGAAAACCTGTCTTTTGAAGGGCTTTGAAAACACGAGG	856
Db	1024	AGAG	CCAGTTCCCTCATCAATGGGAAACCTTTCTATTTCCACGGTGTCAACAAAGATGAGG	1083
QY	857	AAT	CCCCGTTTCGGGGGACGGGACCTTTTATTCATTGATGATATAAAGACTTCAACCTTC	916
Db	1084	ATG	CGGACATCCGAGGGAGGGCTTCGACTGGCCGCTGCTGGTGAAGGACTTCAACCTGC	1143
QY	917	TGA	AGTGGATCAACGCGAATTTTTCAGGACCTCTACTATCTTACAGTGAAGAGTGGC	976
Db	1144	TTG	CGCTGGCTTGGTGGCCACGCTTTCGATCCAGCCACTACCCCTATGCAAGAGAAGTGA	1203
QY	977	TGG	ATCTTGGCCGACAGACTCGGAATCCTTGTGATAGACGAAGCCCCGCA	CGTTGTATCA
Db	1204	TGCA	GATGTGTGACCGCTATGGGATGTGTGTCATCGATGAGTGTCCCGGCGTGGCCCTGG	1263
QY	1037	CAAG	TTACCA---CTACAATCCCGAGACTCAGAAAGATAGCAGAAGACAA	CACTAAGAAGA
Db	1264	CGCT	GGCCGAGTTCTTCAACAACGTTTCTCTGCATCACCACATGCAGGTGATGAAGAAG	1323
QY	1094	TGAT	TCGACAGACACAAGAACCATTCCCAAGTGTGATCATGTGAGTGTGGCAACGAAC	CAG
Db	1324	TGG	TGCGGTAGGACAAAGAACCAACCCCGCGTGTGATGTGTGTGTGGCAACGAGCCTG	1383
QY	1154	AGT	CAACCATCCAGACGCGGAGGGTTCCTTCAAAAGCCCTTTATGAGACTGCCAATGAAA	1213
Db	1384	CGT	CCACCTAGAAATCTGTCTGGCTACTACTTGAAGATGTTGATTCGCTCACACCAAAATCCT	1443
QY	1214	TGG	ATCGAACACAGCCCCGTTGTTCATGTGTAGCATGTATGACGCAACGAGACAGAACA	AA
Db	1444	TGG	ACCCCTCCCGGCTGTGACCTTTGTGAGCACTTAATATGACG-----	1491
QY	1274	GAG	ACGTGGCGCTGAAGTACTTGCACATCGTCTGTGTGAACAGGTACTACGGCTGTACA	1333
Db	1492	CAGA	CAAGGGGCTCCGTATGTGATGTGATCTGTGTGAACAGCTACTACTCTTGTATC	1551
QY	1334	TCT	ATCAGGGAAGATAGAAAGAACTTCAAGCTCTGAAAAAAGACATAGAAGCTCT	1393
Db	1552	ACGA	CTACGGGCACTGGAGTTGATTCAGCTGCAGCTGGCCACCAGTTTGAACACTGT	1611
QY	1394	ATG	CAAGGCACAGAAAGCCCATCTTTGTCA	CAGAAATTGCGTGGCAGCGCATAGCTGCA
Db	1612	ATA	GAAGTATCAGAAAGCCCATTAATTCAGAGCGAGTATGAGCAGAAACGATTCAGAGGT	1671
QY	1454	TCCA	CTACGATCACTCAAAATGTTCTCCGAAGATACCAAGCAGAGCTCGTTGAAAAG-	1512
Db	1672	TTC	ACCAGATCACTCTGTATGTTCAC	TGAAGTACCAGAAAAGTCTGTAGAGCAGT
QY	1513	ACGA	TCAGGCTCCTTTGAAAAA-----AG	ACTATCATCGGAAACACACGCTGTGGCCT
Db	1732	ACCA	TCTGGGTCTGGATCAAAAAACGAGAAAATATGTGTTGAGAGCTCATTTGGAATT	1791
QY	1568	TTG	CAGATTTTAAAGCTCCTCAGAAATGTGAGAAAGCCCATTTCAACCAACAGGGTGT	1627
Db	1792	TTG	CCGATTTTCATGACTGAAACAGTCA	CCGACGAGTGTCTGGGAATMAAAAGGGGATCT
QY	1628	TCACA	AGAGACAGACAACCCAAACTCGTTGCTCATGTACTGAGA	1671
Db	1852	TCAC	TGGCAGAGACAACCAAAAGTGCAGCGCTTCTTTTGCA	1895

RESULT	14
ADJ62864	
ID	ADJ62864 standard; DNA; 2191 BP.
XX	
AC	ADJ62864;
XX	
DT	06-MAY-2004 (first entry)
XX	

DE Human beta-glucuronidase gene related to cancer treatment.
XX
XX beta-catenin; bipartite T-cell factor; Tcf; promoter construct; LEF-1;
KW cytosolic; beta-catenin activity inhibitor; gene therapy; colon cancer;
KW metastasis; liver; thymidine kinase; prodrug; chemotherapy;
KW radiation therapy; surgery; human; beta-glucuronidase; gene; ds.
XX
OS Homo sapiens.
XX
PN US2003228285-A1.
XX
PD 11-DEC-2003.
XX
PF 05-MAY-2003; 2003US-00429802.
XX
PR 03-MAY-2002; 2002US-0377672P.
XX
PA (HUNG/) HUNG M.
PA (KWON/) KWONG K Y.
PA (ZOU/) ZOU Y.
XX
PI Hung M, Kwong KY, Zou Y;
XX
DR WPI; 2004-042209/04.
XX
PT Novel viral vector comprising beta-catenin/bipartite T-cell factor-
PT responsive promoter having first and second promoter region linked to
PT target nucleic acid sequence, useful for treating colon cancer.
XX
PS Disclosure; SEQ ID NO 16; 114pp; English.
XX
XX This invention relates to a novel viral vector comprising a beta-
CC catenin/bipartite T-cell factor (Tcf)-responsive promoter construct which
CC contains a first promoter region having a copy of Tcf/LEF-1 binding site
CC operatively linked to a second promoter region, and a nucleic acid
CC sequence, where the first and second promoter regions are operatively
CC linked to target nucleic acid sequence. The invention may be useful for
CC the development of compounds with a cytosolic activity, through action
CC as beta-catenin activity inhibitors, or for gene therapy. The invention
CC may be useful for treating an individual with colon cancer which is
CC metastasized to the liver. The treatment involves administering the
CC vector of the invention where the nucleic acid sequence encodes a
CC therapeutic polypeptide or thymidine kinase, a prodrug and chemotherapy,
CC radiation, surgery or gene therapy to the individual. The present
CC sequence is that of a therapeutic gene which may be used in the vector of
CC the invention for the treatment of an individual with cancer.
XX
XX Sequence 2191BP; 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;
XX

[illegible]

QY	1037	CAAGGTACCA---CTACAATCCCGAGACTCAGAAAGATAGCAGAAGACAACATTAAGAAGAA	1093
Db	1264	CGCTGCCGACAGTTCTTCACAACACGTTTCTCTGCATCACCACATGCAGGTGATGGAAGAA	1323
QY	1094	TGATTCGACAGACACAAGAACCATCCCAAGTGTGATCATGTGAGGTGGCGAAAGAACAG	1153
Db	1324	TGGTCCGTAGGGGACAAGAACCAACCCCGCGGTGATGATGTGCTGTGGCCAAAGAGCCTG	1383
QY	1154	AGTCCAACCATCCAGACGCGGAGGGTTCTTCAAAGCCCTTTATGAGACTGCCAATGAAA	1213
Db	1384	CGTCCCACTTAGAATCTGCTGGCTACTACTTGAAGATGTGATGCTCACACCAATCCT	1443
QY	1214	TGGAATCGAACACGCCCCGTTGTCTATGGTGAGCATGATGACCGCACAGACGAGAACA	1273
Db	1444	TGGACCCCTCCCGGCTGTGACCTTTGTGAGCACTTAACATAATGCAG-----	1491
QY	1274	GAGACGTGGCGCTGAAGTACTTGCACATCGTCTGTGTGAACAGGTACTACGCGTGTACA	1333
Db	1492	CAGACAAGGGGCTCCGCTATGTGATGTGATCTGTTGAACAGCTACTACTCTTGGTATC	1551
QY	1334	TCTATCAGGGAAGAGTAGAAGAGACTTCAAGCTCTGAAAAAGACATAGAAGAGCTCT	1393
Db	1552	ACGACTACGGGCACTGTGAGTTGATTCAGCTGCAGCTGGCCACCCAGTTTGAGAACTGGT	1611
QY	1394	ATGCAAGGACACAGAAAGCCCATCTTTGTCAACAGAAATTCGTTGGGACGCGAATGACTGCA	1453
Db	1612	ATAAGAAGTATCAGAAAGCCCATTAATTCAGAGCAGATATGAGCAGAAACGATTCAGAGGT	1671
QY	1454	TCCACTACGATCCACCTCAAAATGTTCTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAAG-	1512
Db	1672	TTCACCAGATCCACTCTGATGTTCACCTGAAGAGTACCAAGAAAAGTCTGCTAAGACAGT	1731
QY	1513	ACGATCAGGCTCCTTTTGAAGAA----AGACTACATCATCGGAACACACGCTGTGGCCT	1567
Db	1732	ACCATCTGGGTCTGGATCMAAAACGAGAAAATATGTGTTGGAAGAGCTCATTTGGAATTT	1791
QY	1568	TTGCAGATTTTAAGACTCCTCAGAAATGTGAGAGAACCCATTCTCAACCAACAGGGTGT	1627
Db	1792	TTGCCGATTTTCATGACTGMAACAGTCAACCGACGAGAGTCTGGGGAATMAAAAGGGATCT	1851
QY	1628	TCACAAGAGACAGACAACCCCAACTCGTTGCTCATGTACTGAGA	1671
Db	1852	TCACCTCGGCAGAGACAAACCAAAAGTGACGCTTCTTTTGGCA	1895

```

RESULT 15
ADP10326
ID      ADP10326 standard; DNA; 2191 BP.
XX
XX      ADP10326;
AC
DT      12-AUG-2004 (first entry)
XX
XX      Reference mRNA sequences for marker probe #3.
DE
XX      transplant rejection; immune system; rheumatoid arthritis; lupus;
KW      inflammatory bowel disease; multiple sclerosis; HIV; AIDS; ss.
XX
OS      Homo sapiens.
XX
XX      WO2004042346-A2.
PN
XX
XX      21-MAY-2004.
PD
XX
XX      24-APR-2003; 2003WO-US012946.
PF
XX      24-APR-2002; 2002US-00131831.
PR      20-DEC-2002; 2002US-00325899.
XX
XX      (EXPR-) EXPRESSION DIAGNOSTICS INC.
PA
XX
XX      Wohlgemuth J, Fry K, Woodward R, Ly N, Prentice J, Morris M,
PI      Rosenberg S;

```

XX
DR MPI; 2004-400724/37.
XX
XX
PT Diagnosing or monitoring transplant rejection, e.g. heart, kidney, liver,
PT pancreas, pancreatic islet, lung, bone marrow or stem cell transplant
PT rejection, in an individual, comprises detecting the expression level of
PT the genes.
XX
XX
PS Claim 80; SEQ ID NO 335; 1762pp; English.
XX
XX The present invention relates to diagnosing or monitoring transplant
CC rejection, e.g. cardiac or kidney transplant rejection, in an individual
CC comprises detecting the expression level of one or more genes. The
CC methods, system and kits are useful in diagnosing or monitoring
CC transplant rejection, e.g. heart, kidney, liver, pancreas, pancreatic
CC islet, lung, bone marrow or stem cell transplant rejection,
CC xenotransplant rejection or mechanical organ replacement rejection, in an
CC individual. The method is also useful in assessing the immune status of
CC an individual. The methods are also useful in diagnosing and monitoring
CC diseases that involve the immune system, e.g. rheumatoid arthritis,
CC lupus, inflammatory bowel diseases, multiple sclerosis, HIV/AIDS or
CC viral, bacterial or fungal infection. The present sequence represents a
CC mRNA reference sequence for a 50 mer oligonucleotide marker for diagnosis
CC and monitoring of allograft rejection and other disorders.
XX
XX Sequence 2191 BP; 497 A; 589 C; 630 G; 475 T; 0 U; 0 Other;

Query Match	10.2%;	.Score 172.4;	DB 12;	Length 2191;
Best Local Similarity	52.1%;	Pred. No. 9e-43;		
Matches 492: Conservative	0;	Mismatches 431;	Indels 21;	Gaps 4;

QY	737	TGGAACTTGA AAAAGACGAGTACACTCTGGACATCGGAATCAGAA CGATCAGCTGGAGC	796
Db	964	TGGGGCCTGTGCTGACTTCTACACACTCCCTGTGGGATCGCACTGTGGCTGTACCA	1023
QY	797	AGAAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGGCTTTGAAACGACGAGG	856
Db	1024	AGAGCCAGTTCTCATCAATGGGAAACCTTTCTATTTCCACGGTGTCAACAGCATGAGG	1083
QY	857	AATCCCCGTTCTGGGGCAGGGCACCTTTTATTCATTGATGATPAAAAGACTTCAACCTTC	916
Db	1084	ATCGGACATCCGAGGGAGGGCTTCGACTGGCCGCTGCTGGTGAAGACTTCAACCTGC	1143
QY	917	TGAAGTGATCAACGGAATCTTTCAGAGCCTCTCACTATCCTTACAGTGAAGTGGC	976
Db	1144	TTCCGTGGCTTGGTGC CAACGCTTTCGTACAGCCACTACCCCTATGCAAGAGAA GTGA	1203
QY	977	TGGATCTTGC CGACAGACTCGGAATCCTTGATAGACGAAGCCCGCACGTTGGTATCA	1036
Db	1204	TGCAGATGTGTGACCGCTATGGGATTGTGTGCATCGATGAGTGTCCCGCGGGCCCTGG	1263
QY	1037	CAAGGTACCA---CTACAATCCCGAGACTCAGAAGATAGCAGAAGACA CATAGAAGAA	1093
Db	1264	CGCTGCCGCA GTTCTTCAACAACGTTTCTCTGCATCACCAATGCA GTGATGGAAGAAG	1323
QY	1094	TGATGCACAGACACAAGAACCATCCAGTGTGATCATGTGGAGTGTGGCAACGAACCA G	1153
Db	1324	TGCTGCGTAGGGAACAAGAACCA CCGCGGTGCTGATGTGTGTGTGCGCAACGAGCCTG	1383
QY	1154	AGTCCAACCATCCAGACGCGGAGGGTTCCTTCAAGCCCTTATGAGACTGCGCAATGAAA	1213
Db	1384	CGTCCCACTAGAAATCTGCTGGCTACTACTTGAAGATGTTGATCGCTCACCAAAATCCT	1443
QY	1214	TGGATCGAACACGCCCCGTTGTCTATGTTGAGCATGATGAGCGCACCCAGACGAGAGAACA	1273
Db	1444	TGGACCCCTCCCGGCTGTGACTTGTGTGAGCAACTCTAATCTATGCAG-----	1491
QY	1274	GAGACGTGGCGCTGAAGTACTTGCACATCGTCTGTGTGAACAGTACTACGCTGGTACA	1333
Db	1492	CAGACAAAGGGGGCTCCGTA TGTGATGTGATCTGTTTGAACAGCTACTACTCTTGGTATC	1551
QY	1334	TCTATCAGGGAAGATAGAGAAGGACTTCAAGCTCTGGA AAAAGACATAGAGAAGCTCT	1393

```

Db      1552 ACGACTACGGGACCTGGAGTTGATTCACTGCAGCTGGCCACCAGTTTGAGAACTGGT 1611
QY      1394 ATGCAAGGCACAGAAAGCCCATCTTTGTACAGAAATTCGGTCCGACCGGATAGCTGGCA 1453
      |||||
Db      1612 ATAGAAAGTATCAGAAAGCCCATTTATTCAGAGCGAGTATGGAGCAGAAACGATTGCAAGGT 1671
QY      1454 TCCACTACGATCCACCTCAATGTCTCCGAAAGTACCAAGCAGAGCTCGTTGAAAAG- 1512
      |||||
Db      1672 TTCACCAAGATCCACCTCTGATGTCTCACTGAAGAGTACCAGAAAAGTCTGCTAGACAGT 1731
QY      1513 ACGATCAGGCTCCTTTGAAAAA-----AGACTACATCATCGGAACACACGTTGGGCT 1567
      |||||
Db      1732 ACCATCTGGGCTTGATCAAAAAACGAGAAAATATGTGTTGGAGAGCTCATTTGGAATT 1791
QY      1568 TTGCAGATTTTAAGACTCTCAGATGTGAGAAAGACCCATTCTCAACCAAGGGTGT 1627
      |||||
Db      1792 TTGCCGATTTTCATGACTGAACAGTCAACCGACGAGAGTCTGGGGAAATAAAAAGGGGATCT 1851
QY      1628 TCACAAGAGACAGACAACCCAACTGTTGCTCATGTACTGAGA 1671
      |||||
Db      1852 TCACTCGGCAGAGACAACCAAAAGTGCAAGCGTTCTTTGGCA 1895

```

Search completed: September 1, 2005, 22:25:44
 Job time : 985 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2005, 21:07:30 ; Search time 5837 Seconds
(without alignments)
11014.314 Million cell updates/sec

Title: US-09-936-759-14
Perfect score: 1689
Sequence: 1 atggttaagaccgcaacgaaa.....gaagactgtgagtgaggtc 1689

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hic: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gss1: *
9: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	174	10.3	1377	3	CR593823 full-leng
2	174	10.3	1853	6	CD014093 90135266
3	174	10.3	1995	6	CD014092 90134967
4	161.6	9.6	1124	6	CD503076 CDA60-C07
5	161.2	9.5	1138	6	CD503098 CDA60-D07
6	159.6	9.4	2274	3	AK041058 Mus muscu
7	150.6	8.9	816	7	CK467358 938649 MA
8	144.4	8.5	906	5	BQ941196 AGENCOURT
9	142.6	8.4	878	7	CN155220 942501 MA
10	141.6	8.4	1051	5	BX363460 BX363460
11	132.6	7.9	846	6	CF255373 mdvnl27 c
12	131	7.8	846	6	CA453907 AGENCOURT
13	129.8	7.7	900	6	CA489317 AGENCOURT
14	129.8	7.7	914	6	CB203472 AGENCOURT
15	128.2	7.6	862	7	CO394571 AGENCOURT
16	126.4	7.5	2473	3	AY321342 Rattus no
17	121	7.2	857	5	BX745933 BX745933
18	117.4	7.0	1287	6	CD503099 CDA60-D07
19	117.2	6.9	802	5	BU357212 603474094
20	116	6.9	725	5	BU315788 BM441165
21	114.2	6.8	724	5	BM441165 BM441165
22	112.6	6.7	771	4	BG121498 602352830
23	111.6	6.6	790	7	CO559438 AGENCOURT
24	110.8	6.6	740	7	CF521612 AGENCOURT

25	109.6	6.5	752	4	BU733141 BU733141
26	109.4	6.5	1268	6	CD503077 CDA60-C07
27	107.6	6.4	673	6	CB841568 M15E-2121
28	107.4	6.4	791	5	BU239978 BU239978
29	107.2	6.3	915	5	BQ678153 AGENCOURT
30	106.6	6.3	543	5	BQ360163 QV0-OT003
31	106.6	6.3	615	4	BG568800 602588157
32	106.4	6.3	975	4	BG285429 602406582
33	105.6	6.3	616	2	BE179118 RCO-HT061
34	105.6	6.3	637	1	AJ724809 AJ724809
35	105.4	6.2	504	5	BU917201 EST143 Bo
36	105.2	6.2	736	5	BP454056 BP454056
37	104.6	6.2	757	4	BI851705 603377940
38	104	6.2	877	7	CF406520 CH3#042_G
39	101.8	6.0	1023	5	BX363459 BX363459
40	101.6	6.0	773	7	CF748482 UI-M-HJ0-
41	101	6.0	777	4	BG741172 602631852
42	100.8	6.0	667	7	CN264045 170004243
43	100.6	6.0	548	2	BE751040 202745 MA
44	100.2	5.9	799	5	BU366987 603585124
45	100	5.9	701	5	BX611242 BX611242

ALIGNMENTS

RESULT 1	CR593823	1377 bp	mRNA	linear	HTC 21-JUL-2004
LOCUS	full-length cDNA clone CS0D1001YM21 of B cells (Ramos cell line)				
DEFINITION	Cot 25-normalized of Homo sapiens (human).				
ACCESSION	CR593823.1	GI:50474630			
VERSION	HTC; CNSLT_cDNA.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
REFERENCE	1 (bases 1 to 1377)				
AUTHORS	Li, W.B., Gruber, C., Jessee, J., and Polayes, D.				
TITLE	Full-length cDNA libraries and normalization				
JOURNAL	Unpublished				
REMARK	Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Paraday Avenue				
REFERENCE	2 (bases 1 to 1377)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.				
FEATURES	Location/Qualifiers				
source	1. 1377				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="CS0D1001YM21"				
	/tissue_type="B cells (Ramos cell line) Cot 25-normalized"				
	/plasmid="pCMVSPORT_6"				
ORIGIN					
Query Match	10.3%; Score 174; DB 3; Length 1377;				
Best Local Similarity	52.2%; Pred. No. 7.7e-40;				
Matches	493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;				
QY	737 TGGAACTTGAAGAGAGAGTACCTTGACATCGAATCAGATCAGTGGAGCG				796
Db	215 TGGGCGCTGTGTGACTTCTACACACTCCCTGTGGGAGATCCGACTGTGCTACCA				274

QY 797 AGAAGGCTCTATCTGAACGGAACTGTCTTTTGAAGGCTTGGAAAGCAGAG 856
 Db 275 AGAGCAGTTCCTCATCATGGAACTTTCTATTTCCAGGTGCAACAGCATGAGG 334
 QY 857 AATTCCCGTTCTGGGCGAGGCACTTTTATCCATTGATGATTAAGAATTCACTTC 916
 Db 335 ATGCGGACATCCGAGGAAAGGCTTGCACTGGCCGCTGCTGTGAAGGACTTCACTGC 394
 QY 917 TGAAGTGATCAACGGCAATTCTTGAAGACCTTCACTATCTTACATGAAAGTGGC 976
 Db 395 TTGCTGGCTTGTGTCGAACGCTTTCGTACCAAGCACTACCTATGACAGAAAGTGA 454
 QY 977 TGGATCTTGGCAGACACTCGGAATCTTGTGATAGCAAGCCCGCACTTGGTATCA 1036
 Db 455 TGCAAGTGTGACCCCTATGGATGTGTGATCATGATGATGTCCTCCGCGCTGG 514
 QY 1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAAAGATAGCAGAAACAACATAAGAGAA 1093
 Db 515 CGTGCCGAGTTCCTTCAACACGTTTCTTGATCACCACATGAGGTGATGAAAGAG 574
 QY 1094 TGATCGACAGACAGAAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGCAACAG 1153
 Db 575 TGTGCGGTAGGACAGAAACACCCCGGCTGTGATGTGTGCTGTGGCCAAACGAGCTG 634
 QY 1154 AGTCCAACCATCCAGACGCGGAGGTTTCTTCAAGCCCTTATGAGACTGCCAATGAAA 1213
 Db 635 CGTCCACCTAGAAATGTGCTGCTACTTGAAGATGGTGTGATCGCTCACACCAATCCT 694
 QY 1214 TGGATCGAACACGCGCCGTTGTCAATGAGCATGATGAGCGACAGACAGAGAAACAA 1273
 Db 695 TGGACCCCTCCCGCTGTGACCTTGTGAGCAACTTAATATGAG----- 742
 QY 1274 GAGACGTGGCGCTGAATCTTCAAGTGTGCTGTGGAACAGGTACTACGCTGTGACA 1333
 Db 743 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTGAACAGTACTACTTGTGATC 802
 QY 1334 TCTATCAGGGAAGATAGAAGAGACTTCAAGCTCGAAAAAGACATAGAAGACTCT 1393
 Db 803 ACGACTACGGGCACTGGAAGTGTGATCAGCTGACGCTGGCCACCAAGTTGAGAACTGGT 862
 QY 1394 ATGCAAGGCACAGAAAGCCCATCTTGTGTACAGAAATTCGCTGCGGACGCGATAGCTGCA 1453
 Db 863 ATAGAAGTATCAGAAGCCCATTAATCAGAAGCATATGAGCAGAAACGATTCGAGGGT 922
 QY 1454 TCCACTACGATCCACCTCAATGTCTCCGAAGATACCAAGCAGAGCTCGTTGAAAG- 1512
 Db 923 TTCAACAGGATCCACTCTGATGTTCACTGAAGATACCAAGAAAGTCTGCTAGACAGT 982
 QY 1513 ACGATCAGGCTCCTTTGAAAAA-----AGACTACATCATCGGAACAACGTTGGGCT 1567
 Db 983 ACCATCTGGGTCTGATCAAAAAACGAGAAATACGTGTTGAGAGCTCATTTGGAATT 1042
 QY 1568 TTGAGATTTTAAGACTCCTCAGAAATGTGAGAAGAACCATTTCAACCAAGGGGTTT 1627
 Db 1043 TTGCCGATTTTCATGACTGAACAGTCAACGACGAGAGTCTGGGGAATAAAAAGGGATCT 1102
 QY 1628 TCACAAGAGACAGAACCAACCAACTGCTGCTCATGTAAGTGA 1671
 Db 1103 TCACTCGGACAGAGCAACCAAAAAGTGCAGCGTTCTTTTGGCA 1146

RESULT 2
 CD014093 1853 bp mRNA linear EST 21-OCT-2003
 LOCUS 90135266 Single gene library Homo sapiens cDNA, mRNA sequence.
 DEFINITION CD014093
 ACCESSION CD014093
 VERSION CD014093.1 GI:37777622
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 1853)
 AUTHORS Jin,P., Fu,G.K., Wilson,A.D., Yang,J., Chien,D., Hawkins,P.R.,
 TITLE Au-Young,J. and Stuve,L.L.
 JOURNAL PCR isolation and cloning of novel splice variant mRNAs from known
 COMMENT drug target genes
 CONTACT Genomics 83 (4), 566-571 (2004)
 CONTACT Contact: Jin, P.
 INCYTE Incyte Corporation
 3160 Porter Drive, Palo Alto, CA 94304, USA
 TEL: 650 621 8639
 FAX: 650 621 8965
 EMAIL: pjn@incyte.com.
 FEATURES Location/Qualifiers
 source 1..1853
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="Single gene library"
 /note="Vector: pDrive Cloning Vector; RT-PCR was performed using gene-specific primers flanking the open-reading frame. PCR products were subcloned into pDrive Cloning Vector and sequenced completely using M13 forward and reverse primers. Sequencing gaps were closed by re-sequencing using primers flanking the gapped areas."
 ORIGIN
 Query Match 10.3%; Score 174; DB 6; Length 1853;
 Best Local Similarity 52.2%; Pred. No. 8.5e-40;
 Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;
 QY 737 TGGAACTTGAAGAAAGACGAGTACACTGTGACATCGGAATGAAACGATCAGCTGGAGC 796
 Db 624 TGGGGCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGCTGCACCA 683
 QY 797 AGAAGAGGCTTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAG 856
 Db 684 AGAGCAGTTCCTCATCAATGGGAAACCTTCTATTCCACGGTGTCAACAGCATGAGG 743
 QY 857 AATTCCCGTTCTGGGGCAGGGCACTTTATTCATTGATGATTAAGAGCTTCAACCTTC 916
 Db 744 ATGCGACATCCGAGGGAAGGGCTTGCATGCGCCGCTGTGTGAAGACTTCAACCTGC 803
 QY 917 TGAAGTGATCAACGCGAATCTTTCAGAGCTCTCACTATCTTACAGTGAAGTGGC 976
 Db 804 TTGCTGCTGTGGTCCAAAGCTTTCCTGACAGCCACTACCTCTATGACAGGAAGTGA 863
 QY 977 TGGATCTTGGCGACAGACTCGGAATCTTGTGATGAGCAGAAAGCCCGCAGTGGTATCA 1036
 Db 864 TGCAAGTGTGACCGCTATGGGAATGTGTGATGATGATGATGATGATGATGATGATGAT 923
 QY 1037 CAAGGTACCA---CTACAATCCGAGACTCAGAAAGATAGCAGAAACAACATAAGAGAA 1093
 Db 924 CGCTGCCGAGTCTTCAACAAAGTTTCTGTGATCACCACATGACAGGTGATGAAAGAG 983
 QY 1094 TGATCGACAGACAGAAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGCAACAG 1153
 Db 984 TGTGCGTAGGAGCAAGAACCAACCCGCGGTGTGATGTGTGTGTGTGTGTGTGTGTGTGT 1043
 QY 1154 AGTCCAACCATCCAGACGCGGAGGGTTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAA 1213
 Db 1044 CGTCCACCTAGAATCTGTGCTACTACTGAAGATGTGATGCTCACACCAAAATCCT 1103
 QY 1214 TGGATGAAACAGCCCGCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1273
 Db 1104 TGAACCCCTCCCGCTGTGACCTTTGTGAGCAACTTAATATGACAG----- 1151
 QY 1274 GAGACGTGGCGCTGAAGTACTTTCAGACATCGTGTGTGAAACAGTACTACGCTGTACA 1333
 Db 1152 CAGACAAAGGGGCTCCGTATGTGATGTGATGTGATGTGATGTGATGTGATGTGATGTGAT 1211
 QY 1334 TCTATCAGGGAAGATAGAAGAGACTTCAAGCTCTGGAAGAAAGACATAGAAGACTCT 1393
 Db 1212 ACGACTACGGGCACTGGAATGATTCAAGCTGACGCTGCGCACCCAGTTTGAGAACTGCT 1271

preceded by a synthetic XhoI site. 5 prime adaptors were used containing an EcoRI cohesive end. The finished cDNAs were inserted in to the ZAP express vector unidirectionally in the sense orientation with respect to the lacZ promoter of PBK-CMV. An amplified library was prepared from approximately 3 million primary clones in the lambda ZAP Express vector. In vivo excision was then used to generate individual PBK-CMV phagemid clones for EST sequencing."

ORIGIN

Query Match	9.5%;	Score 161.2;	DB 6;	Length 1138;
Best Local Similarity	51.9%;	Pred. No. 4.7e-36;		
Matches	473;	Conservative 0;	Mismatches 418;	Indels 21; Gaps 4;
OY	748 AAAGACGAGTACACTCGGACATCGGAATCAGACGATCAGTGGGACGAGAAGAGCTC 807			
DB	71 AAGCAGGTGTACACTCTACACAGTCGGCATCCGCACGGTCGACGTTACCAAGCAGCTGATATT 130			
OY	808 TATCTGAACGGAAACCTGTCTTTTGAAGGGCTTTGGAAGAAGCAGGGAATTCCTCCGTT 867			
DB	131 CTGATCAACAAAAAGCCCTTCTACTTCCACGGAGTAATAACACGAGACTCTGATATT 190			
OY	868 CTGGGCGAGGGCACCCTTTTATCCATTGATGATAAAAGACTTCAACCTTCTGAAGTGATC 927			
DB	191 CGAGGCAAAAGGCGCTGAGCTGGCCCCCTCATGTGTGAAGACTTTAACTTATGAAGTGGTTG 250			
OY	928 AACGGCAATTCTTTCAGAGACCTCTACTATCTTACAGTGAAGAGTGGCTGGATCTTGGC 987			
DB	251 GGGGCCAACTCGTCCGCACCAAGCCACTACCTTATGACAGAGAGATCCTGCAGATGTGT 310			
OY	988 GACAGACTCGGAATCCTTGTGATAGAGAAAGCCCGCAGCTTGGTATCACA--AGGTAC 1044			
DB	311 GACCGCCATGGCATCGTGTGTGATAGCAGAGTGCCCCGGCGTGCGCATAAAGACATTCGC 370			
OY	1045 CACTACAATCCCGAGACTCAGAAAGATAGCAGAAACATTAAGAAGATGATCGACAGA 1104			
DB	371 AGTTTGGAAAACGCCCTCTTAACCCCATCACTGTGTCATGACGAGCTGTACGTGG 430			
OY	1105 CACAAGAACCATCCCAAGTGTGATCATGTGGAGTGTGGCGAAGCAAGACAGAGTCCAAACCAT 1164			
DB	431 GACAAAGAACCATCCCTCTGTGTGTCATGTGTCAAGTGCCCAATGAGCCGGCTGCAGAGATG 490			
OY	1165 CCAGACGCGGAGGGTTTCTTCAAAAGCCCTTATGAGACTGCCAATGAATGATCGACACA 1224			
DB	491 CCTCCTGTGATTACTATTTCAAAACCTTGATAAACATACCAAAAGATTGGATCCAAAC 550			
OY	1225 CGCCCGGTGTGATGAGCATGATGAGCGCACAGACGAGAGAAAGAGACGTGGCG 1284			
DB	551 CGGCCCGT-----CACTTTTATCAGACAGATACTATGCGAGGATAAAGGG 598			
OY	1285 CTGAAGTACTTTCGACATCGTCTGTGTGAACAGGTAATAAGGCTGTGATCATCTATCAGGGA 1344			
DB	599 GCTCCCTACGTGACGTCATCTGCGTAAACAGTTACTTCTCCTGTACCATGACCCGGGC 658			
OY	1345 AGGATGAAGAAGAGACTTCAAGCTTGGAAAAAAGACATAGAAGAGCTTATGCAAGGCAC 1404			
DB	659 CACCCGAGGTCACTCCCATCCAGCTCAACACTCAGTTGAGAACTGTGTACGGAAGTAC 718			
OY	1405 AGAAAGCCCATCTTTGTACAGAAATTCGGTGGCGGCGATAGTGGCATCCACTACGAT 1464			
DB	719 CAGAAACCCATCATCAAGAGCGAATACGAGCGGATGCGTGCCGGGGCTTACAGATGAT 778			
OY	1465 CCACCTCAATGTTCTCCGAGAGTACCAAGCAGAGCTCGTTGAAAAGACGATCAGGCTC 1524			
DB	779 CCACCCGTGATGTTTACTGAGAGTACCAAGATTAGTCTGACAGGCTTACCAACGTCG 838			
OY	1525 CTTTGAAGAAAAG-----ACTACATCATCGGAACACACGTTGGGCTTTGAGATTTTA 1579			
DB	839 TTGCACCAGAAAAGACAGTACGTATCGGGAACCTCATCTGAACTTTGACAGACTTCA 898			
OY	1580 AGACTCCTCAGAAATGGAAGAACCATTTCTCAACCAAGGAGTGTTCACAAAGAGACA 1639			

DB	899 TGACCACACAAAGGATCATGCG-CGTGTTGGGAAACAAGAGGTGTCTTCAGCAGAGAAA 957
OY	1640 GACAACCCCAAC 1651
DB	958 GGACGCCCAAGC 969

RESULT 6

AK041058

LOCUS

DEFINITION

AK041058 2274 bp mRNA linear HTC 03-APR-2004
Mus musculus adult male aorta and vein cDNA, RIKEN full-length enriched library, clone:A530072005 product:beta-glucuronidase structural, full insert sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AK041058 GI:26088231
HTC; CAP trapper.
Mus musculus (house mouse)

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2274)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoaka, T., Hirokane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshinide Hayashizaki, The Institute of

OY 1425 AGAATTCGGTCCGACGCGATAGCTGGCATCCACTACGATCCACCTCAATGTTCTCCGA 1484
DB 589 CGAGTATGAGCAGAAACGATTTGACGGTTTACACGAGATCCACCTCTGATGTTCACTGA 648
OY 1485 AGAGTACCAAGCAGAGCTCGTTGAAAG-ACGATCAGGCTCCTTTGAAAAAG-----A 1538
DB 649 AGAGTACCAGAAAAAGTCTGCTAGAGCAGTACCATCTGGTCTGGATCAAAAAACGAGAAA 708
OY 1539 CTACATCATCGGAACACACGTTGGCCCTTTGCAGATTTTAAGACTCTCTCAGAAATGTGAG 1598
DB 709 ATACGTGTTGAGAGCTCATTTGGAATTTGCCGATTTTCATGATGAACAGTCAACCGAC 768
OY 1599 AAGA-CCCATTTCTCAACCAAGGGTGTTCACAAGACAGACAGAACCCAAACTCGTTG 1657
DB 769 GAGAGTGTGGGGGAATAAAGGGGATCTTCACCTCGCAGACAGAACCAAAAAAGTGACAG 828
OY 1658 CTCATGTACTGAGAAGACTGTGAGTGAAGTT 1689
DB 829 CGTTCCTTTTGGCAGAGAGATAGTGGAGATT 860

RESULT 9
CN155220
LOCUS CN155220 878 bp mRNA linear EST 02-APR-2004
DEFINITION 942501 MARC 4PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION CN155220
VERSION CN155220.1 GI:46169650
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa

REFERENCE
AUTHORS Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 878)
Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
Nonneman,D.J., Wray,J.E. and Keele,J.W.
Porcine EST collection using a normalized library constructed from
embryos representing early developmental stages
Unpublished (2003)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Baases called with phred v0.020425 c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: TMW8046 row: I column: 19
Seq primer: GTAATACGACTCATTATAGG.

JOURNAL
COMMENT
Unpublished (2003)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Baases called with phred v0.020425 c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: TMW8046 row: I column: 19
Seq primer: GTAATACGACTCATTATAGG.

FEATURES
Source
1. 878
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tisue_type="pooled"
/lab_host="DH10B"
/clone_id="MARC 4PIG"
/note="vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;
library made with combined RNA from day-10, day-13,
day-15, day-25, and day-30 whole embryos."
ORIGIN

Query Match 8.4%; Score 142.6; DB 7; Length 878;
Best Local Similarity 52.1%; Pred. No. 1.5e-30;
Matches 425; Conservative 0; Mismatches 369; Indels 21; Gaps 4;

OY 866 TTCTGGGGGACGACCTTTTATCCATGATGATATAAAGACTTCAACTTCTGAAGTGA 925
DB 6 TCCGAGGCAAGGGCTTTGACTGTCCTGCTGTGTAAGGACTTCAACTGCTCCGCTGGC 65
OY 926 TCAACGCGAATCTTTCAAGACCTCTCACTATCCTTAAGTGAAGTGGCTGATCTTG 985

DB 66 TGGGTGCCAAGCCTTCCGACCAAGCCACTACCCCTACGACAGAGAGGTGATGCAGCTCT 125
OY 986 CCGACAGACTCGGAATCCTTGTGTAGACGAAGCCCCGACGTTGTATACAAAGTACC 1045
DB 126 GTGACCGTTATGGGATCGTGTGATCATGATGAGAGTCCGGGTGTGGCATCGTCTGGCCC 185
OY 1046 A---CTACATCCCGAGACTCAGAAGATAGCAAGAACAATAAGAAGATGATCGACA 1102
DB 186 AGAGCTTCAGAACGCGCTCTCTGCAGACACCACTAGAGGTGATGAGAGATGCTTCGCA 245
OY 1103 GACACAAGAACCATCCCAAGTGTGATCATGTGGAAGTGTGGCGAACGAACCAAGTCCAAC 1162
DB 246 GGGACAAGAACCAACCCGCGGTGTGATGTGTCCTGGCCAAAGAGCTTCTCTCC 305
OY 1163 ATCCAGACGCGGAGGGTTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAATGATCGAA 1222
DB 306 TGGAACAGGCTGCTTACTTACTTCAAGATGCTGATCGGCCACCAAGCCTTGACCCCT 365
OY 1223 CACGCCCCGTTGTATGTGATGATGATGACCGCACCAAGAGAGAGAACAGACAGCTGG 1282
DB 366 CCGGCCCCGTGACCTTCTGACCA-----GCTCAGCTATGA----AAAAGACCTGG 413
OY 1283 CGCTGAAGTACTTCGACATCGTCTGTGTGAACAAGGTACTACGCGCTGTATCTATCAG 1342
DB 414 GGGTGCCGTATGTGAGCGTGAATCTGTGTGAACAAGTACTCTCTGTATCATGACTACG 473
OY 1343 GAAGGATAGAAGAGGACTTCAAGCTCTGGAAGAAAGACATAGAAAGAGCTCTATGCAAGGC 1402
DB 474 GTCACATGAGGTGATTCAGCTGCAGCTGGCCACCAAGTTGAGAGGTGGCATGAGGCGCT 533
OY 1403 ACAGAAAGCCCATCTTGTGACAGAAATTGGGTGGAGCGGACGATAGCTGGCATCCACTACG 1462
DB 534 ACCAGAAACCAATTATTCAGAGCGAGTACGGGGCAGAAACCATCATAGGGTTTCACGAGG 593
OY 1463 ATCCACCTCAAATGTTCTCCGAGAGTACCAAGCAGAGCTGTGAAAAGAGATCAGGC 1522
DB 594 ATCCACCTCATGTTGTTGAGTGAAGATACAGAAAGGCGCTGCTACAGCAGTATCATGTGA 653
OY 1523 TCCTTTTGAAA-----AAAGACTCATCATCGGAACACAGCTGTGGCCCTTTGCAAGATT 1576
DB 654 TTCTGATCAAAAACGCAAGAAATATGTGTTGGAAGCTGATCTGAAATTTTGGCCGATT 713
OY 1577 TTAAGACTCCTCAGAATGTGAGAAAGACCATTTCTAACCCACAGGGGTGTTTCAACAAGAG 1636
DB 714 TTATGACCGCAATCATCACTCAGAGAGGGCCATAGGGAATAGAAAAGGATCTTCACTGAC 773
OY 1637 ACAGACAACCAAACTCGTTGCTCATGTACTGAGA 1671
DB 774 AGAGACAACCAAAAAGTGACAGCATTCCTGTTGCGA 808

RESULT 10
BX363460
LOCUS BX363460 1051 bp mRNA linear EST 08-APR-2004
DEFINITION BX363460 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
Homo sapiens cDNA clone CS0DL001YM21 5-PRIME, mRNA sequence.
ACCESSION BX363460
VERSION BX363460.2 GI:46291531
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1051)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 5, 2003 this sequence version replaced gi:30372720.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 4169.r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?S=CS0DL001AG11QPI&c=4169.r.

FEATURES

Source

```

1..1051
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL001YM21"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

```

ORIGIN

Query Match 8.4%; Score 141.6; DB 5; Length 1051;
Best Local Similarity 50.7%; Pred. No. 3.2e-30;
Matches 412; Conservative 5; Mismatches 379; Indels 16; Gaps 3;

```

QY 737 TGGAACTTGAAAAAGCAGTACACTTGACATCGGAATCAGAACATCAGCTGGGACG 796
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 215 TGGGGCGCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGACTGTGCTGTACCA 274

QY 797 AGAAGAGGCTCTATCTGAACGGGAACTGTCTTTTGAAGGGCTTTGGAAGCAGCAGG 856
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 275 AGAGCCAGTTCCTCATCAATGGGAACTTTCTATTTCCACGGGTGTCAACAAGCATGAGG 334

QY 857 AATTCCTGCTTGTGGGCAAGGACCTTTTATCCATGATGATAAAGACTTCAACCTTC 916
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 ATGCGACATCCGAGGGAAGGGCTTGCAGCTGCGCGCTGTGTGAAGACTTCAACCTGC 394

QY 917 TGAAGTGGATCAACGGGAATTCTTTAGAGACCTCTCATCTATCCTTACAGTGAAGTGGC 976
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 395 TTCGCTGGCTTGTGTGCAACGCTTTCGTACAGCCACTACCCCTATGACAGGAAGTGA 454

QY 977 TGGATCTTCCGACAGACTCGGAATCTTGTGATAGCAAGCCCGCAGCTTGTATCA 1036
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 455 TGCAGATGTGTGACCGCTATGGAGTGTGTGATCGATGAGTGTCCGGCGTGGCGCTGG 514

QY 1037 CAAGGTACCA--CTACAATCCCGAGACTCAGAAGATAGCAGAAACAATAGAGAA 1093
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 515 CGTGCAGCAGTCTTCAACAACGTTTCTGTGATCACCATGCAAGTGAAGAAAG 574

QY 1094 TGATCGACAGACAGAACAACCATCCAGTGTGATCATGTGAGTGTGGAGCAAGCAAG 1153
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 575 TGTGTGCTAGGAGCAAGAACCACCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 634

QY 1154 AGTCCAACCATCCAGACGCGGAGGTTTCTTCAAAAGCCCTTATGAGACTGCCAATGAAA 1213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 CGTCCACCTAGAAATCTGTGCTACTACTTGAAGATGTGATCGCTACACCAAAATCCT 694

QY 1214 TGGATCGAACACAGCCCGTGTGTATGATGAGCATGATGAGCAGCAGCAGAGAGACAA 1273
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 695 TGGACCCCTCCCGCGCTGTGACCTTGTGAGCAACTCTAATATGACG----- 742

QY 1274 GAGAGTGGCGCTGAAGTACTTTCAGATCTGTGTGTAACAGGTACTACGGCTGTACA 1333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 743 CAGACAAGGGGCTCCGTATGTGATGTGATCTGTGTAACAGCTACTACTTGTGTATC 802

QY 1334 TCTATCAGGGAAGATAGAAAGAGCTTCAAGCTTGAAAAAAGACATAGAAAGCTCT 1393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 803 ACGACTACGGGACCTGGAAGTGTGATTCAGCTGACGTGGCCACCCAGTTTGAGAACTGGT 862

QY 1394 ATGCAAGGCACAGAAAGCCCATCTTGTTCACAGAAATTCGCTGCGGAGCGGATA-GCTGGC 1452
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 863 ATAAGAAGTATCAGAACCCCATTAATTCAGAGCCGAGTATGAGCAGAGAAAGATTTCAGGGG 922
QY 1453 ATCACTACGATCCACTCAATGTTCTCCGAAGATACCAAGCAGAGCTGTTGAAAAG 1512
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 923 TTTCACAGAGATCCACTCTGTATGTTCAGTGAAGATTAACAGAAATTTTCTAGACAG 982

QY 1513 ACGATCAGGCTCTTTTGAAGAAAGACTACAT 1544
Db 983 TACCATCTGGTCTGTGATTAAGAAAGGAAAT 1014

```

RESULT 11

CF255373

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

CF255373 767 bp mRNA linear EST 07-AUG-2003
mdm127 c10 r2 Marek's disease virus-infected spleen Gallus gallus
CDNA, mRNA sequence.
CF255373
CF255373.1 GI:33488628
EST.
Gallus gallus (chicken)
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
Witzell, H., Bed'Hom, B., Morin, V., Young, J.R., Whittaker, C.J.,
Chausee, A.M. and Zoorob, R.
A collection of chicken ESTs from activated immune cells
Unpublished (2003)
Contact: Zoorob R
UPR 1983

COMMENT

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

COMMENT

FEATURES

Source

```

1..767
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone_lib="Marek's disease virus-infected spleen"
/note="Organ: Spleen; Vector: pTriplex2"

```

ORIGIN

Query Match 7.9%; Score 132.6; DB 7; Length 767;
Best Local Similarity 50.5%; Pred. No. 1.4e-27;
Matches 386; Conservative 0; Mismatches 364; Indels 15; Gaps 2;

```

QY 751 GACGATACACTTGAACATCGGAATCAGAAGATCAGCTGGACGAGAAGGCTCTAT 810
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 GACATCTACACACTGCCGTTGGCATCCGACCGTGACAGTACACAGCAGCAGATTCCTC 61

QY 811 CTGAACGGGAACCTGTCTTTTGAAGGCTTTGGAAGCAGCAGGAATTCCTCGTCTG 870
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 ATCAACGGAAGCCCTTCTACTTCCACGGGTCAACAAGCAGAGACGCGATATTCGT 121

QY 871 GGGCAGGACACTTTTATCCATGATGATAAAGACTTCAACCTTGAAGTGAATCAAC 930
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 GGCAGAGGCTGTGACTGGCAGTATGTTAAGACTTCAACCTGCTGCTGTTGGGG 181

QY 931 GCGAATTTCTTCAGACCTCTCACTATCCTTAAGTGAAGAGTGTGATCTTGGCGAC 990
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 GCAAACTCTTCGTACAGCACTACCCCTATGTGAGAGATCATGACCTGTGTGAC 241

QY 991 AGACTCGAATCTTGTGATAGCAAGCCCGCAGCTTGTGTATCAAGG--TACCAC 1047
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 242 ACCTACGGCATTTGTGTATGACAGAGTGCCCGGCGTGGGATTAAGATGCTGAGAGC 301

QY 1048 TACAATCCGAGACTCAGAAGATAGCAAGACACATAAAGAAATGATGACAGACAC 1107
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 TTGGGAAACAATCTCTACAGCATCATCTGCTGTGATGAGGAAGTATCCGAGGAT 361

```


OY		1108	AAGAACCATCCCACTGTCATGTGATCATGTGGAGTGTGGCGAACGAACCAAGAGTCCAAACCATCCA	1167
Db		362	AAGAACAAGGCCCTCAGTTGTGATGTGGTCAAGTACCAATGAGCCGGGCATCAGAGCTGCC	421
OY		1168	GACCGGAGGGTTTCTTCCAAAGCCCCCTTATGAGACTGCCAATGAATGGATCGAACAACGC	1227
Db		422	CCAGCAGCTTACTACTTCAAGACACTGATAGCTCACACTAAAGCTCTGGATCCCTCCAGA	481
OY		1228	CCCCGTGTCATGGTGAGCATGATGATGACGCACCACGACGAGAGAACAGAAGACGTCGGCTG	1287
Db		482	CCTGTAACGTTTGTGTCTGATACTAATTACGCTGC-----GATCATGGTGCT	529
OY		1288	AAGTACTTCGACATCGCTGTGTGNAACAGGTACTACGGCTGTATCATCTATCAGGGAAG	1347
Db		530	CCTTATGTGATGATAATCTGTGTGAACAGCTACTTCTCCTGGTACCAAGACCGGCCAT	589
OY		1348	ATAGAAGAAGACTTCAAAGCTCTGAAAAAAGACATAGAAGAGCTTATGCAAGGCACAGA	1407
Db		590	CTGGAAGTTATTCCACTGCAACTTACAGCACAGTTTGAAACTGGTATTAANAACCTACAG	649
OY		1408	AAGCCCATCTTTGTCAACAGAAATTCGGTGCAGACGCGATAGCTGGCATCCACTACGATCCA	1467
Db		650	AAACCCATTATCCAGAGTAGAATATGAGAGCTGACTCGATTCCCGCACTTCACAGTGATCCA	709
OY		1468	CTCAAAATGTTCTCCGAAGAGTACCAAGCAGAGCTGTTGAAAAG	1512
Db		710	CTGCTATGTTCACTGAGAGAGTATCAGAAAGCTATGCTGAAAGAG	754

RESULT 12	
CA453907	
LOCUS	
DEFINITION	CA453907 846 bp mRNA
	AGENCOURT_10738557 MAPCL Homo sapiens CDNA clone IMAGE:6718553 5', linear EST 12-NOV-2002
mRNA sequence.	

ACCESSION	CA453907
VERSION	CA453907.1
	GI:24903124

KEYWORDS	EST.
SOURCE	Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 846)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT **Contact:** Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Kristi A. Egland, Ira

CDNA Library Preparation: Invitrogen Corp

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.

http://image.lim.gov

Plate: LLAML42/5 ROW: a COLUMN: 1
High quality commercial stock: 708

FEATURES

REVISED
EDITION

```

000000      L : 030
/organiam="Homo sapiens"
```

```

/moj type="mRNA"
/obj name="mRNA"

```

/db xref="taxon:9606"

/cjone="IMAGE:6718553"

/cell line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231

hTERT-HME1, LNCaP"

```
/lab host="EMDH10B"
```

```
/clone lib="MAPCL"
```

/note="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;

Subtracted with brain, liver, lung, kidney and muscle.

Directionally cloned. Priming method: oligo-dT. Average

insert size: 1800 bp. Library amplification: 26,000 fold.

Kristi A. Eglyand, James J. Vincent, Robert Strausberg,

Bungkook Lee & Ira Pastan: Discovery of new breast

cancer genes encoding membrane and secreted proteins.

Manuscript submitted."

ORIGIN	Query Match	7.8%;	Score 131;	DB 6;	Length 846;
	Best Local Similarity	52.4%;	Pred. No. 4.4e-27;		
	Matches 417;	Conservative	0;	Mismatches 356;	Indels 23; Gaps 5;
QY	866 TTCTGGGGCAGGGCACCTTTTATCCATTGATGATAAAGACTTCAACCTTGTGAAGTGA				925
Db	61 TCCGAGGGAAAGGGCTTCGACTGGCCGCTGCTGGTGAAGAGACTTCAACCTTGCCTGGC				120
QY	926 TCAACGGCAATCTTTTCAAGAACCTCTCACTATCCTTACAGTGAAGAGTGGCTGATCTTG				985
Db	121 TTGTGCCAAAGCTTTCCGTACCAAGCCACTACCTTATGCAAGAGAAAGTATGCAGATGT				180
QY	986 CCGACAGACTCGGAATCCTTGTGATAGACGAAGCCCGCAGCTGTGTATCAAGGTACC				1045
Db	181 GTGACCGCTATGGGATTTGGTTCATGATGATGTCCCGCGCTGGCGCTGCCGC				240
QY	1046 A---CTACAATCCGAGACTCAGAAAGATAGCAAGAACAACATAAGAAATGATGCACA				1102
Db	241 AGTTCCTTCAACAAAGTTTCTCTGCATCACCACATGCAGGTGATGAAAGATGTGCGTA				300
QY	1103 GACACAAGAACCATCCCAAGTGTATCATGTGAGTGTGGCGAAGCAACAGAGTCCAACC				1162
Db	301 GGGACAAGAACCAACCCCGCGGTGCTGATGTGTCTGTGGCCAAAGAGCTGCTCCACC				360
QY	1163 ATCCAGACCGCGGAGGGTTCTTTCAAGCCCTTTATGAGACTGCCAATGAATGATCGAA				1222
Db	361 TAGAATCTGCTGGCTACTACTTGAAGATGTGATCGCTCAACCAATCCTTGAACCCCT				420
QY	1223 CACGCCCCGTTGTCTATGTTGATGACATGATGACGCCACCAAGCAGAGAAACAAGAGCGTG				1282
Db	421 CCGGCGCTGTGACCTTTGTGAGCAACTCTAATATGCAG-----CAGACAAG				468
QY	1283 CGGTGAAGTACTTCGACATCGTCTGTGTGAACAGGTACTACGGCTGTACATCTATCAGG				1342
Db	469 GGGCTCCGTATGTGATGTGATCTGTTTGAACAGCTACTTCTTGGTATCAGACTACG				528
QY	1343 GAAAGATAGAAGAAGAGCTTCAAGCTCTGAAAAAAGACATAGAAGAGCTCTATGCAAGGC				1402
Db	529 GGCACCTGAGTTGATTACAGCTGCAGCTGGCCACCAAGTTGAGAAGCTGTTAAGAAGT				588
QY	1403 ACAGAAAGCCCATCTTTGTCA CAGAATTCGGTGGCGGACGGATAGCTGGCATCCACTACG				1462
Db	589 ATCAGAAGCCCATTTATTCAAGGCGAGTATGAGGACAGAAACGATTTGCAGGGTTTCAACGAG				648
QY	1463 ATCCACCTCAAAATGTTCTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAG--ACGATCAGG				1521
Db	649 ATCCACCTCTGATGTTCACTGAAGAGTACCAAGAAAGTCTGCTAGAGCAGTACCATCTGG				708
QY	1522 CTCCTTTGAAAAAAGAC-----TACATCATCGGAACACAGTGTGGCTTTGCAGAT				1575
Db	709 GNTCTGATCAAAAAAGCAGAAAAATACGTGGTTGGAGAGCTCATTTGGAATTTGCCGAT				768
QY	1576 TTTAAGACTCTCAGAAATGTGAAAGACCCCAATCTCAA--CCACAAGGGTGTTTTCACAAG				1634
Db	769 TTTATGACTGAACAGTCAACCGACGAGAGTGTGGGAATTAAAAAAGGGATCTTCACTCG				828
QY	1635 AGACAGACAACCCAAA 1650				
Db	829 GCAGAGACCACCCAAA 844				

RESULT 13	CA489317	LOCUS	DEFINITION	CA489317	900 bp	mRNA	linear	EST 14-NOV-2002
				AGGCCOURT_10810259	MAPcL	Homo sapiens	CDNA clone IMAGE:6721700 5',	
				mRNA sequence.				

ACCESSION CA489317

VERSION CA489317.1 GI:24952108

KEYWORDS EST.

SOURCE	Homo sapiens (human)
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Homo sapiens	1 (bases 1 to 900)	NIH-MGC	http://mgs.nci.nih.gov/	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
CONTACT: Robert Strausberg, Ph.D.					
EMAIL: cgapbs-r@mail.nih.gov					
TISSUE Procurement: Kristi A. Eglund, Ira Pastan					
CDNA Library Preparation: Invitrogen Corp					
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)					
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:					
http://image.lnl.gov					
Plate: LLM14283 row: d column: 20					
High quality sequence stop: 580.					
FEATURES					
source					
1.900					
/organism="Homo sapiens"					
/mol_type="mRNA"					
/db_xref="taxon:9606"					
/clone="IMAGE:6721700"					
/cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231, hTERT-HME1, LNCaP"					
/lab_host="EMDH10B"					
/clone_11b="MAPcl"					
/notes="Vector: pCMV-SPORT6; Site_1: EcoRV; Site_2: Not I; Subtracted with brain, liver, lung, kidney and muscle. directionally cloned. Priming method: oligo-dT. Average insert size: 1800 bp. Library amplification: 26,000 fold. Kristi A. Eglund, James J. Vincent, Robert Strausberg, Bungkok Lee & Ira Pastan: Discovery of new breast cancer genes encoding membrane and secreted proteins. Manuscript submitted."					
ORIGIN					
Query Match	7.7%	Score 129.8;	DB 6;	Length 900;	
Best Local Similarity	51.9%	Pred. No. 1e-26;			
Matches 402;	Conservative 0;	Mismatches 352;	Indels 21;	Gaps 4;	
866	TTCTGGGGCAGGCGACCTTTTATCCATTGATGATPAAAGACTTCAACCTTCTGAAGTGA	925			
31	TCCGAGGGAAGGCGCTTGCAGTCGCCGCTGCTGCTGAAGACTTCAACCTTCTGCTGGC	90			
926	TCAACGCGAATCTTTCAGGACCTCTCACTATCCTTACAGTGAAGAGTGGCTGATCTTG	985			
91	TTGTGTCACAGCGCTTCCGTACAGCCACTACCCCTATGACAGAGAGATGATGATGT	150			
986	CCGACAGACTCGGAATCCTTGTGATAGACGAAGCCCCGACGCTTGTATCAAGGTACC	1045			
151	GTCACCGCTATGGGATTTGTGTCATGATGAGTGTCCGCGCTGGGCTGGCGTCCGC	210			
1046	A---CTACAATCCGAGACTCAGAAAGATAGCAGAAGACAACATTAAGAAGATGATGACA	1102			
211	AGTCTTCAACAAGCTTTCTCTGCATCACACATGACAGTGTGAGAGAGTGTGCGTA	270			
1103	GACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGCAACAGATCCCAAC	1162			
271	GGGACAAGAACCAACCCGCGGTGTGATGTGTGTGGCCAAGAGCTGTGCCACC	330			
1163	ATCCAGACGCGAGGGTTTCTTCAAAAGCCCTTATGAGACTGCCAATGAATGATGAA	1222			
331	TAGAACTGTGCTGCTACTACTTGAAGATGTGATCGCTCACACCAATCTTGGACCCCT	390			
1223	CACGCCCCGTTGTATGTGAGCATGTAGACGACACAGACGAGAGAACAGAGAGCTGG	1282			
391	CCCGGCTGTGACCTTGTGAGCAACTTAATATGACG-----CAGACAAGG	438			
1283	CGCTGAAGTACTTGACATCGTCTGTGGAACAGTACTACGCGCTGTATCTATCAGG	1342			
439	GCGCTCCGATGTGATGTGATCTGTTTGAACAGCTACTACTTGTGTATCAGACTACG	498			

Query	1343	GAAGGATAGAAAGGAGCTTCAAGCTCTGTGAAAAAAGACATAGAAAGAGCTTATGCAAGGC	1402
Db	499	GGCACCTGGAGTTGATTACAGCTGCAGCTGGCCACCCAGTTTGAGAACTGGTATTAAGAGT	558
Qy	1403	ACAGAAAGCCCATCTTTGTGTCTCAGAAATTGGGTGGAGCGGATAGCTGGCATCCACTACG	1462
Db	559	ATCAGAAAGCCCATTTATTACAGAGCGAGTATGGAGCAGAAACGATTTCAGGGTTTCACCAGG	618
Qy	1463	ATCCACCTCAAAATGTTCTCCGAGAGTACCAGACGAGCTCGTTGAAAAG-ACGATCAGG	1521
Db	619	ATCCACCTCTGATGTTCACTGAGAGTACCAGAAAAGTCTGCTAGAGCAGTACCATCTGG	678
Qy	1522	CTCCCTTTGAAAAAAG-----ACTACATCATCTGGAAACACACGTGTGGCCCTTGCAGATT	1576
Db	679	GTCCTGATCAAAAACGACGAAAATACGTGGTTGGAGAGCTCATTTGAAAATTTGCCGATT	738
Qy	1577	TTAAGACTCCTCAGAAATGTGAGAAAGCCCATTTCTCAACCAGAGGCTGTTTTCAC	1631
Db	739	TCATGACTGAACGGTCCACCGACGAGAGTGTCTGGGGAATATAAAAGGGGATCCTC	793
RESULT 14			
CB203472		914 bp	mRNA
LOCUS			linear
DEFINITION	CB203472	AGENCOURT_11297304 NIH_MGC_135 Mus musculus cDNA clone	EST 05-FEB-2003
ACCESSION	CB203472	IMAGE:30141354 5', mRNA sequence.	
VERSION	CB203472.1	GI:28239894	
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 914)		
AUTHORS	NIH-MGC http://mgi.nci.nih.gov/.		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@rs-r@mail.nih.gov Tissue Procurement: Dr. David Rowe cDNA Library Preparation: Invitrogen Corp cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: NDAM0048 row: n column: 19 High quality sequence stop: 580.		
FEATURES			
source			
	1. 914		
	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:30141354"		
	/lab_host="NIH MGC 135"		
	/note="Vector: pCMVSPORT6.1; Site_1: EcoRV; Site_2: NotI; Normalized full-length enriched library from pooled mouse embryonic limb, maxilla and mandible, day 12.5, 13.5, 14.5, and 15.5 (size selected for the 0.5-1 kb fragments) cloned directionally, priming method: Oligo-dT. cDNA enrichment: >1k bp, Average insert size 1.6k bp. Normalization (Cot value): 7.5 kb. Priming sequence: 5'/GAC TAGTTC TAGATCGCGAGCGGCCGCC(T)3' Tissue contributed by, David Rowe. Library constructed by Resgen, Invitrogen Corp."		
ORIGIN			
Query Match	7.7%;	Score 129.8;	DB 6; Length 914;
Best Local Similarity	50.4%;	Pred. No. 1e-26;	
Matches 410;	Conservative 0;	Mismatches 382;	Indels 21; Gaps 3;

QY 871 GGGCAGGGCACCCTTTATCCATTGATGATAAAGACTTCAACCTTCTGAAGTGATCAAC 930
DB 3 GGGAAAGGCTTCGACTGGCCGCTGCTGGTAAAGATTTCACCTGCTCCGTGGCTCGGG 62
QY 931 GCCAATCTTTCAGAGACCTCTCACTATCTTACAGTGAAGAGTGGCTGATCTTGCCGAC 990
DB 63 GCAATTCCTTTCGATCCAGCCATATCCCTACTCAGAGGAGGTACTTCACTCTGTGAC 122
QY 991 AGACTCGGAATCCTGTGATAGACGAAGCCCGCAGCTGG---TATCACAAGTACCAC 1047
DB 123 CGATACGGGATTGTGTCATCGATGATGTCCCGGTGTGGCATTTGCTACCTCAGAGT 182
QY 1048 TACAATCCCGAGACTCAGAAGATGACAGAACAACATAGAAGATGATCGACAGACAC 1107
DB 183 TTTGGCAACGAGTCACTTCGGCACCACTAGAGGTGATGAGAGCTGGTCCGCGGGAC 242
QY 1108 AAGAACCATCCAGTGTGATCATGTGAGAGTGTGGCAACGAACCAAGTCCAACCATCCA 1167
DB 243 AAAAAATCACCTCGCGGTGTGATGTGCTGTGGCCAATGAGCCTTCCTCTGCTTGAAA 302
QY 1168 GACGCGGAGGGTTTCTTCAAGCCCTTTATGAGACTGCCAATGAATGGATCGAACAGCC 1227
DB 303 CCGCGCCGATATTACTTTAAGACGCTGATCACCCACACCAAGCCCTGGACCTCACCCGT 362
QY 1228 CCGGTGTGATGTCAGCATGATGACGACGACGACGAGAGAAACAAGACGTCGCGCTG 1287
DB 363 CCGGTGACCTTTGTGAGCA-----ACGCCAATATGATGACAGACCTGGGGGCC 410
QY 1288 AAGTACTTCGACATGCTGTGTGACAAGGTACTACGGCTGTGATCTTATCAGGGAAGG 1347
DB 411 CCGTACGTGATGATGTTATCTGTGTAACAGCTACTTTTCTGTGATCATGACTATGGGCAT 470
QY 1348 ATAGAAGAAGGACTTCAAGCTCTGGAAGAAAGACATAGAAGGCTCTATGCAAGGCACAGA 1407
DB 471 TTGGAGGTGATTACGCCACAGCTGAATAGCCAGTTGAGAAGCTGGTATTAAGACGCATCAG 530
QY 1408 AAGCCCATCTTTGTGACAGAAATTCGGTGGGACGCGATAGCTGGCATCCACTAGCATCCA 1467
DB 531 AAGCCGATTATCCAGAGCGAGTATGAGACAGACCAATCCAGGGATCCACGAGACCCG 590
QY 1468 CCTCAATGTTCTCCGAAGAGTACCAAGCAGAGCTCGTGA---AGACGATCAGG 1521
DB 591 CCTCGCATGTTCACTGAGAGAGTACCAAGGCTGTTCTGAGAATTACCATTCAGTTCTG 650
QY 1522 CTCTTTTGAAGAAAGACTACATCATTCGGAACACAGCTGTGGCCTTTGACAGATTTAAG 1581
DB 651 GATCAGAAAAGTAAGAAATACGTGGTGGAGAGCTCATCTGAATTTGCGCGACTTCATG 710
QY 1582 ACTCCTCAGAAATGAGAGAAGACCACTTCTCAACCAAGGGGTGTTTCAAGAGACAGA 1641
DB 711 ACGAACCACTACCGCTGAGAGTAATCGGGAACAAGGGGATCTTCACTCGCAGAGA 770
QY 1642 CAACCCAAACTCGTTGCTCATGTACTGAGAGA 1674
DB 771 CAGCCCAAAACTCGGCTTTATTTTTCGAGA 803

RESULT 15
LOCUS CO394571 862 bp mRNA linear EST 01-JUL-2004
DEFINITION AGENCOURT 27822917 NIH_MGC 252 Rattus norvegicus cDNA clone
IMAGE:7307606 5', mRNA sequence.
ACCESSION CO394571
VERSION CO394571.1 GI:49576487
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 862)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics / NIH
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: Drs. Josef Lazar & Howard Jacob, Medical
College of Wisconsin
cDNA Library Preparation: Express Genomics
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM15349 row: e column: 12
High quality sequence stop: 685.

FEATURES
source

1. 862
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="IMAGE:7307606"
/tissue_type="Ovary, strain - Norway line3. Age8 weeks.
Tissue was snap-frozen adn transferred in -70C. RNase free
for the entire procedure"
/lab_host="DH10B Tona"
/clone_lib="NIH_MGC_252"
/note="Organ: ovary; Vector: pDONR 201; Site_1: NotI;
Site_2: NotI; RNA obtained from female ovaries animals at
8 wk old. Tissues were snap-frozen and kept at -80C for
two days before RNA extraction and purification
(Tri-reagent method). cDNA was primed using oligo-dT
primer: 5'-pGACTAGTCTAGATCGGAGCGCGCCGCT(T)25-3' and
cloned into the pScorV/NotI sites of pExpress-1.
Size selection >1.25kb resulted in an average insert size
of 1.7kb. This primary library is not normalized
(normalized library is NIH_MGC_252) and was constructed by
Express Genomics (Frederick, MD). Note: this is a NIH_MGC
library"

ORIGIN

Query Match 7.6%; Score 128.2; DB 7; Length 862;
Best Local Similarity 51.5%; Pred. No. 3e-26;
Matches 354; Conservative 0; Mismatches 318; Indels 15; Gaps 2;
QY 855 GGAATCCCGCTCTGGGGCAGGCGACCTTTATCCATTGATGATAAAGACTTCAACCT 914
DB 7 GGATTCGGGATCGAGGAGAGGCTTCGACTGGCCTCTGCTGATAAAGATTCAACCT 66
QY 915 TCTGAAGTGATCAACGCGAATTTCTTCAGAGCCTCTCACTATCCTTACAGTGAAGAGTG 974
DB 67 CTTCCGTTGGCTCGGGCAAAATTCCTTCGTACCAAGCCACTATCCCTACTCGAGAGAGGT 126
QY 975 GCTGATCTTGGCCGACAGACTCGGAATCCTGTGATAGACGAAGCCCGCAGCTTGCTAT 1034
DB 127 ACTCAGCTCTGTGACCGATATGAATGTGTGATCATCGATGAGTGTCCCGGTGTGGCAT 186
QY 1035 CACAAGTACCA---CTACAATCCCGAGACTCAGAAGATAGAGAAGAACAAACATAAGAG 1091
DB 187 CGTGTGCCCCAGAGTTTGGCAACGTGTCTCTTCGGCACCACTTAGAGGTGATGACGA 246
QY 1092 AATGATCGACAGACAGAAGAACCATCCAGTGTGATCATGTGGAAGTGTGGCCAAAGAAC 1151
DB 247 GCTGTGCGCAGGAGCAAAAATCAACCTGCGGTGTGATGTGTGTGTGCGCAATGAGCC 306
QY 1152 AGAGTCCAACCATCCAGACGCGGAGGGTTTCTTCAAGCCCTTTATGAGACTGCCAATGA 1211
DB 307 TGTCTCTTCTTGAAACCTGCGGATATTACTTCAAGACGCTGATCGCCACACCAAGC 366
QY 1212 AATGATCGAACACGCGCGCTGTGATGTCATGTCATGATGAGCGCACACGAGAGAGAAC 1271
DB 367 CTTGACCCCAACCCGCTCTGTGACCTTTGTGAGCAATA-----CCAGATATGA 414

```

QY      1272 AAGAGACGTGGCCGCTGAAAGTACTTCCGATCGTCTGTGTGAACAGGTAAGTACTACGGCTGGTA 1331
          |||||
Db      415 CGCAGACCTGGGGGCCCCCGTACGTGAGTGATTTGTGTGAACAGTTACTTATCCTGGTA 474
          |||||
QY      1332 CATCTATCAGGGAAGGATAGAGAAGGACTTCAAGCTCTGAAAAAGACATAGAGAGCT 1391
          |||||
Db      475 TCATGACTACGGGCATCTGGAGGTGATTCAGCTGCAGCTGACTAGCCAGTTGAGAACTG 534
          |||||
QY      1392 CTATGCAAGGCACAGAAAGCCCATCTTTGTCCACAGAAFTCGGTGCGGACGGATAGCTGG 1451
          |||||
Db      535 GTATAGATGTACACAGAGCCCAATTATCCAGAGCGAGTATGAGCAGACGCCGCTCGGG 594
          |||||
QY      1452 CATCCACTACGATCCAAATGTCTCCGAAGAGTACCAGCAGAGCTCGTGAAA 1511
          |||||
Db      595 GCTTCATGAGGATCCACTCGCATGTTCAGTGAGAGTACACAGACGCTCTCTGGAGAA 654
          |||||
QY      1512 GACGATCAGGCTCCTTTGAAAAAGA 1538
          |||||
Db      655 TATCATTTGATCCTGGATGAGAAACGA 681
          |||||

```

Search completed: September 2, 2005, 02:08:04
 Job time : 5847 secs

This Page Blank (uspto)


```
QY 361 AAAGTGGAGAGATTTCCTCGAAGTTCAGACAGCGGCACTCACACCGTGGATTATTTT 420
Db 361 AAAGTGGAGAGATTTCCTCGAAGTTCAGACAGCGGCACTCACACCGTGGATTATTTT 420
QY 421 GGAAGTTTTCCACCTGCAAACTTCGACTTCTTCCCTACGCTGAATCATTAAGCCTGTT 480
Db 421 GGAAGTTTTCCACCTGCAAACTTCGACTTCTTCCCTACGCTGAATCATTAAGCCTGTT 480
QY 481 CTGATGAGTTTCACAGACCAGCGAGGATACTCGACATCTGGGTGACACGAGTGAAGTCT 540
Db 481 CTGATGAGTTTCACAGACCAGCGAGGATACTCGACATCTGGGTGACACGAGTGAAGTCT 540
QY 541 GAAACCGGAGAGAACTTGAAAAAGTGAAGTAGAAGTAGAAGTCTCAGAAGAGCGGTG 600
Db 541 GAAACCGGAGAGAACTTGAAAAAGTGAAGTAGAAGTAGAAGTCTCAGAAGAGCGGTG 600
QY 601 GGACAGAGAGATGACGATCAAACCTTGAGAGAGAGAAAGATTAGAACATCCACACAGA 660
Db 601 GGACAGAGAGATGACGATCAAACCTTGAGAGAGAGAAAGATTAGAACATCCACACAGA 660
QY 661 TTCTGCGAAGGGGAGTTTCATCTCGAAAAACCGCAGTTCTGAGCCTCGAAGATCCATAT 720
Db 661 TTCTGCGAAGGGGAGTTTCATCTCGAAAAACCGCAGTTCTGAGCCTCGAAGATCCATAT 720
QY 721 CTTTATCCTCTCAAGGTGGAACCTTGAAAAAGACGAGTACACTCTGACATCGGAATCAGA 780
Db 721 CTTTATCCTCTCAAGGTGGAACCTTGAAAAAGACGAGTACACTCTGACATCGGAATCAGA 780
QY 781 ACGATACGCTGGGACGAGAAAGGCTTACTTGAACGGGAAACCTTCTTTTGAAGGGC 840
Db 781 ACGATACGCTGGGACGAGAAAGGCTTACTTGAACGGGAAACCTTCTTTTGAAGGGC 840
QY 841 TTTGGAAGCACAGAGAAATTCCTGGGGCAGGGCACTTTTATCCATTGATGATA 900
Db 841 TTTGGAAGCACAGAGAAATTCCTGGGGCAGGGCACTTTTATCCATTGATGATA 900
QY 901 AAAGACTTCAACCTTCTGAAAGTGATGATCAACGCGAATCTTTCAAGACCTCTCACTATCT 960
Db 901 AAAGACTTCAACCTTCTGAAAGTGATGATCAACGCGAATCTTTCAAGACCTCTCACTATCT 960
QY 961 TACAGTGAAGAGTGGCTGGATCTTGCCGACAGACTCGGAATCTTGTGATAGACGAAGCC 1020
Db 961 TACAGTGAAGAGTGGCTGGATCTTGCCGACAGACTCGGAATCTTGTGATAGACGAAGCC 1020
QY 1021 CCGCACGTTGGTATCACAGAGGTACCACCTACATCCCGAGACTCAGAAAGATAGCAGAGAC 1080
Db 1021 CCGCACGTTGGTATCACAGAGGTACCACCTACATCCCGAGACTCAGAAAGATAGCAGAGAC 1080
QY 1081 AACATTAAGAAAGATGATGACAGACACAGAAACCATCCAGTGTGATCATGTGAGTGTG 1140
Db 1081 AACATTAAGAAAGATGATGACAGACACAGAAACCATCCAGTGTGATCATGTGAGTGTG 1140
QY 1141 GCGAACGAACCAAGATCAACCATCCAGACGCGAGGGTTTCTTCAAGCCCTTTATGAG 1200
Db 1141 GCGAACGAACCAAGATCAACCATCCAGACGCGAGGGTTTCTTCAAGCCCTTTATGAG 1200
QY 1201 ACTGCCAATGAAGTGAATGCAACACGCCCCGTGTGATGTTGAGCATGATGAGACGCCA 1260
Db 1201 ACTGCCAATGAAGTGAATGCAACACGCCCCGTGTGATGTTGAGCATGATGAGACGCCA 1260
QY 1261 GACGAGAGAACAAGAGCGTGGCGTGAAGTACTTGCATGCTGTGTGAACAGGTAC 1320
Db 1261 GACGAGAGAACAAGAGCGTGGCGTGAAGTACTTGCATGCTGTGTGAACAGGTAC 1320
QY 1321 TACGGCTGTGATCATCTATCAGGGAAGGATAGAAGAGACTTCAAGCTCTGAAAAAGAC 1380
Db 1321 TACGGCTGTGATCATCTATCAGGGAAGGATAGAAGAGACTTCAAGCTCTGAAAAAGAC 1380
QY 1381 ATAGAAAGCTCTATGCAAGGCACAGAAAGCCCATCTTGTACAGAAATTCGGTGGGAC 1440
Db 1381 ATAGAAAGCTCTATGCAAGGCACAGAAAGCCCATCTTGTACAGAAATTCGGTGGGAC 1440
```

```
QY 1441 GCGATAGCTGSCATCCACTACGATCCACTCTCAATGTTCCTCCGAAGATACCAAGCAGAG 1500
Db 1441 GCGATAGCTGSCATCCACTACGATCCACTCTCAATGTTCCTCCGAAGATACCAAGCAGAG 1500
QY 1501 CTCGTTGAAAAAGCAGATCAGGCTCCTTTTGAAAAAAGACTACATCATCGGAACACAGCTG 1560
Db 1501 CTCGTTGAAAAAGCAGATCAGGCTCCTTTTGAAAAAAGACTACATCATCGGAACACAGCTG 1560
QY 1561 TGGGCTTTGCGAGATTTTAAAGACTCCTCAGAAATGTGAGAAAGACCCATTCTCAACCAAG 1620
Db 1561 TGGGCTTTGCGAGATTTTAAAGACTCCTCAGAAATGTGAGAAAGACCCATTCTCAACCAAG 1620
QY 1621 GGTGTTTCAACAAGACAGACCAACCCAACTCGTTGCTGATGTACTGAGAAAGACTGTGG 1680
Db 1621 GGTGTTTCAACAAGACAGACCAACCCAACTCGTTGCTGATGTACTGAGAAAGACTGTGG 1680
QY 1681 AGTGAGGTT 1689
Db 1681 AGTGAGGTT 1689
```

```
RESULT 2
US-09-715-858-1
; Sequence 1, Application US/09715858
; Patent No. 6582692
; GENERAL INFORMATION:
; APPLICANT: Podsakoff, Gregory
; APPLICANT: Watson, Gordon
; APPLICANT: Couto, Linda B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE
; FILE REFERENCE: 0800-0021
; CURRENT APPLICATION NUMBER: US/09/715,858
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1956
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1956)
US-09-715-858-1
```

```
Query Match 10.2%; Score 172.4; DB 4; Length 1956;
Best Local Similarity 52.1%; Pred. No. 5.5e-43;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

QY 737 TGAAGCTTGAAGAAAGCAGATCACTCTGACATCGGAATCAGAAAGATCAGCTGGAGC 796
Db 938 TGGGGCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGACTGTGCTGTACCA 997
QY 797 AGAAGAGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAG 856
Db 998 AGAGCCAGTTCCTCATCAATGGGAAACCTTCTATTCCACGGTGTCAACAAAGCATGAG 1057
QY 857 AATTCCTGCTGGGAGGCACTTTTATCCATTGATGATTAAGACTTCAACCTTC 916
Db 1058 ATGCGGACATCCGAGGAAAGGCTTGACTGGCCGCTGTGGAAGAGACTTCAACCTGC 1117
QY 917 TGAAGTGAATCAACGGAATCTTTCAGAGCTCTCACTATCTTACAGTGAAGAGTGC 976
Db 1118 TTCCTGTGCTGTGCTGCAACGCTTTCGTACCAAGCCACTACCCCTATGACAGAGAGTGA 1177
QY 977 TGAATCTTGCAGACAGACTCGGAATCTTGTGATAGAGAAAGCCCGCAGCTGTGATCA 1036
Db 1178 TGCAGATGTGTACCGCTATGGGATTTGTGTATCGATGATGATGTCCCGGCTGGCTGG 1237
QY 1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAAAGATAGCAGAAAGACACATAGAAGAA 1093
Db 1238 CGCTGCCGCAATTCTTCAACAACGTTTCTGTGCATCACCAATGCAGAGTGTAGAAAG 1297
```

```

QY 1094 TGAATCGACAGACAGAACCAATCCAGTGTGATCATGTGAGTGTGGCGAAGCAACGAG 1153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1298 TGGTGCCTAGGACAGAACCAACCCCGCGGTGATGTGTCTGTGGCCAAACGAGCCTG 1357
QY 1154 AGTCCCAACCATCCAGACGCGAGGGTTCTTCAAGCCCTTATGAGACTGCCAATGAAA 1213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1358 CGTCCACCTAGAAATCTGTGCTGCTACTACTTGAAGATGGTGTATCGCTCACACCAATCCT 1417
QY 1214 TGGATCGAACACGCCCCGTTGTCTAGTGTGAGCATGATGAGACGACACGAGAGAACAA 1273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1418 TGAACCCCTCCCGGCTGTGACCTTGTGAGCAACTTAATATGAG----- 1465
QY 1274 GAGACGTGGCGCTGAAGTACTTGCATCTGTGTGAGACAGGTAACGAGTACGCGTGTACA 1333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1466 CAGACAGAGGGGGCTCCGTATGTGATGTGATCTGTTGAACAGCTACTCTTGTATC 1525
QY 1334 TCTATCAGGGAAGATAGAGAAGACTTCAAGCTTGGAAAAAGACATAGAGACTCT 1393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1526 ACGACTACGGGCACTGGAGTGTGATTCAGCTGCAGCTGGCCACCCAGTTGAGAACTGT 1585
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTGTACAGAAATTCGGTCCGACGCGATAGCTGGCA 1453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1586 ATAGAAGTATCAGAAAGCCCATTAATCAGAGCGAGTATGAGACAGAAACGATTTGACGGGT 1645
QY 1454 TCCACTACGATCCCACTCAATGTCTCCGAAAGATACCAAGCAGACTCGTTGAAAAAG- 1512
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1646 TTCACCAAGATCCACTCTGTATGTCTCACTGAAGAGTACCAGAAAAAGTCTGCTAGACAGT 1705
QY 1513 ACGATCAGGCTCCTTTTGAAGAA-----AGACTACATCATCGGAACACAGTGTGGCCT 1567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1706 ACCATCTGGGTCTGATCAAAAAACGAGAAATATGTGTTGAGAGCTCATTTGGAAAT 1765
QY 1568 TTGCAAGATTTTAAGACTCTCAGAAATGTGAGAAAGACCATTTCTCAACCAAGGGTGT 1627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1766 TTGCCGATTTTCATGACTGAACAGTCAACGACGAGAGTGTGGGAAATAAAAAAGGGATCT 1825
QY 1628 TCACAGAGACAGACAAACCAACTCGTTGCTCATGTACTGAGA 1671
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1826 TCACTCGGACAGACAAACCAAAAGTGACGCTTCTTTTGGCA 1869

```

RESULT 3

```

US-09-039-555B-12
; Sequence 12, Application US/09039555B
; Patent No. 6033856
; GENERAL INFORMATION:
; APPLICANT: Koerner, Kathrin
; APPLICANT: Mueller, Rolf
; APPLICANT: Sadlacek, Hans-Harald
; TITLE OF INVENTION: PROMOTER OF THE CDC25B GENE, ITS
; TITLE OF INVENTION: PREPARATION AND USE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/039, 555B
; FILING DATE: 16-MAR-1998
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 19710643.9
; FILING DATE: 14-MAR-1997
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 016779/0131
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2191 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-039-555B-12

```

```

Query Match 10.2%; Score 172.4; DB 3; Length 2191;
Best Local Similarity 52.1%; Pred. No. 5.9e-43;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

```

```

QY 737 TGAACCTTGAAGAAAGACAGTACACTCTGACATCGGAATCAGAACGATCAGCTGGGACG 796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 964 TGGGGCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGACTGTGGCTGTACCA 1023
QY 797 AGAAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAG 856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1024 AGAGCAGTTCCTCATCAATGGGAAACCTTCTATTTCCACGGTGTCAACAAAGATGAG 1083
QY 857 AATTCCCGTTCTGGGGCAGGGACCTTTTATCCATTGATGATTAAGACTTCAACTTC 916
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1084 ATGCGACATCCAGGAGGAGGCTTCGACTGGCGCTGTGTGAAGACTTCAACTTGC 1143
QY 917 TGAAGTGAATCAACGCAATTTCTTACAGACCTCTCACTATCTTACAGTGAAGTGGC 976
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1144 TTGCTGGCTTGTGCAACGCTTCCGTACCACTACCTATGACAGAGAACTGA 1203
QY 977 TGGATCTTGGCCAGACTCGGAATCCTTGTGATGAGAAAGCCCGCAGCTGTGATCA 1036
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1204 TGCAGATGTGTACCGCTATGGAATGTGTGATCAGTGAAGTGTCCGCGTGGCCTGG 1263
QY 1037 CAAGGTACCA---CTACAATCCGAGACTCAGAAATAGCAGAAACAACATAGAAGAA 1093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1264 CGCTGCCGAGTTCTTCAACAACGTTTCTGTCATCACACATGCAAGTGATGAAGAA 1323
QY 1094 TGAATCGACAGACAGAACCAATCCAGTGTGATCATGTGAGTGTGGGAACGACAG 1153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1324 TGGTGCCTAGGACAGAACCAACCCCGCGGTGATGTGTGTGTGGCCAAACGAGCCTG 1383
QY 1154 AGTCCCAACCATCCAGACGCGAGGGTTCTTCAAGCCCTTATGAGACTGCCAATGAAA 1213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1384 CGTCCACCTAGAAATCTGTGCTACTACTTGAAGATGTGATCGCTCACACCAATCCT 1443
QY 1214 TGGATCGAACACGCCCCGTTGTCTAGTGTGAGCATGATGAGCGCACGACGAGAACAA 1273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1444 TGAACCCCTCCCGGCTGTGACCTTTGTGAGCAACTTAATATGAG----- 1491
QY 1274 GAGACGTGGCGCTGAAGTACTTGCATCTGTGTGTAACAGTACTACGCGTGTACA 1333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1492 CAGACAGGGGGCTCCGTATGTGATGTGATCTGTTGAACAGTACTACTCTTGTATC 1551
QY 1334 TCTATCAGGGAAGATAGAGAAGACTTCAAGCTCTGAAAAAGACATAGAGACTCT 1393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1552 ACGACTACGGGCACTGGAAGTTCAGCTGCAGCTGGCCACCAAGTTGAGAACTGT 1611
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTGTCTACAGAAATTCGGTCCGACGCGATAGCTGCA 1453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1612 ATAGAAGTATCAGAAAGCCCATTAATTCAGAGCGAGTATGAGCAGAAACGATTTGACGGGT 1671
QY 1454 TCCACTACGATCCCACTCAATGTCTCCGAAAGTACCAAGCAGAGCTCGTTGAAAAAG- 1512
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1672 TTCACCAAGATCCACTCTGATGTTCACGTAAGAGTACCAAGAAAGTCTGCTAGACAGT 1731
QY 1513 ACGATCAGGCTCCTTTTGAAGAA-----AGACTACATCATCGGAACACAGCTGTGGCCT 1567

```

```
Db 1732 ACCATCTGGGCTCTGATCAAAAAACGAGAAAAATATGTGTTGAGAGCTCATTTGGAATT 1791
QY 1568 TTGCAGATTTTAAAGTCTCTCAGAAATGTAGAGAGACCATTTCTCAACCAAGGGTGT 1627
Db 1792 TTGCCGATTTTCATGACTGAACAGTCAACGACGAGAGTGTGGGAATAAAAAGGGATCT 1851
QY 1628 TCACAAGAGACAGACAAACCAACTCGTTGCTCATGTACTGAGA 1671
Db 1852 TCACTCGGCGAGAGACAAACCAAAAGTGCAGCGTTCTTTTGGCA 1895
```

```
RESULT 4
US-09-715-858-3
; Sequence 3, Application US/09715858
; Patent No. 6582692
; GENERAL INFORMATION:
; APPLICANT: Podsakoff, Gregory
; APPLICANT: Watson, Gordon
; APPLICANT: Couto, Linda B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE
; FILE REFERENCE: TREATMENT OF LYSOSOMAL DISORDERS
; CURRENT APPLICATION NUMBER: US/09/715,858
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1947
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1947)
; US-09-715-858-3
```

```
Query Match 9.4%; Score 159.6; DB 4; Length 1947;
Best Local Similarity 50.6%; Pred. No. 5.6e-39;
Matches 477; Conservative 0; Mismatches 444; Indels 21; Gaps 3;
```

```
QY 739 GAACCTGAAAAAGACGAGTACACTCTGACATCGGAATCAGACGATCAGCTGGACGAG 798
Db 928 GAGTCTGTGACTGACTACTACACCCCTTCTATCGGATTCGACAGTGGCTGTCAAAAG 987
QY 799 AAGAGCTCTATCTGAACGGGAACCTGTCTTTTGAAGGCTTTGGAAGACACGAGAA 858
Db 988 AGCAAGTTCCTCATAAACGGGAAGCCCTTCTATTTCGAAGGGTCAATAAGACAGAGAT 1047
QY 859 TTCCCGCTTCTGGGGGACGACCTTTTATCCATTGATGATAAAGACTTCAACCTTCTG 918
Db 1048 TCAGATATCCGAGGGAAGGCTTCGACTGGCCGCTGCTGTAAGAGATTTCACCTGCTC 1107
QY 919 AAGTGAATCAACGCGAATTTCTTACAGACCTCTCACTATCTTACAGTGAAGAGTGGCTG 978
Db 1108 CGTTGGCTCGGGGCAAAATTCCTTTCGTAACGACCACTATCCCTACTCAGAGAGGTACTT 1167
QY 979 GATCTTCCGACAGACTCGGAATCTTGTGATAGACGAAGCCCGGACGTTGG--TATC 1035
Db 1168 CAGCTCTGTGACCGATACGGGATGTGTGTCATCGATGAGTGTCCCGGTGTGGGCAATTGTG 1227
QY 1036 ACAAGGTACCACTACATCCCGAGACTCAGAGATAGCAGAAGACAAACATAAGAAATG 1095
Db 1228 CTACCTCAGAGTTTGGCAACGAGTCACTTCGGACCACTTAGAGGTGATGAGAGCTG 1287
QY 1096 ATCGACAGACACAAGAACCATCCCAAGTGTGATCATGTGAGAGTGTGGCAACGACAGAG 1155
Db 1288 GTTCGCGGGGACAAAATCAACCTCGCGTGTGTGATGTGTCTGTGGCCAATGAGCCTTCC 1347
QY 1156 TCCAACCATCCAGACGCGGAGGGTCTTCAAAAGCCCTTTATGAGACTGCCAATGAATG 1215
Db 1348 TCTGCTCTGAAACCCGCGCATATTACTTTAAGACGCTGATCACCACACCAAGCCCTG 1407
```

```
QY 1216 GATCGAACACGCCCCGTTGTCTATGTGTGAGCATGATGAGCCGACACGAGAGAACAAAGA 1275
Db 1408 GACCTCACCCGTCCTCGTGACCTTTGTGAGCA-----ACGCCAAATATGATGCA 1455
QY 1276 GACGTGGCGCTGAAGTACTTTCACATCGTCTGTGTGAACAGAGTACTACGGCTGGTACATC 1335
Db 1456 GACCTGGGGGCCCCGTACGTGATGTTATCTGTGTAAACAGCTACTTTCTTGGTATCAT 1515
QY 1336 TATCAGGGAAGATAGAGAAGGACTTCAAGCTCTGAAAAAAGACATAGAAGACTCTAT 1395
Db 1516 GACTATGGGCATTTGAGGTGATTCAGCCACGACTGAATAGCCAGTTTGAAACTGGTAT 1575
QY 1396 GCAGGCGACAGAAAGCCCATCTTTGTCAAGAAATTGCGTCCGACCGGATAGCTGGCATC 1455
Db 1576 AAGACGATCAGAAAGCCGATTAATCCAGAGCGAGTATGAGACAGACGCAATCCAGGGATC 1635
QY 1456 CACTACGATCCACTCMAATGTTCTCCGAAGAGTACCAAGCAGAGCTGTTGAAAAAGAGC 1515
Db 1636 CACGAGGACCCGCTCGCATGTTCAGTGAGGAGTACCAAGAGCTGTTCTGGAGAAATTAC 1695
QY 1516 ATCAGGCTCTTTGAAA-----AAAGACTACATCATCGGAACACACGCTGGGCCCTT 1569
Db 1696 CATTCAGTTCTGATCAGAAACGTAAAGAAATACGTGTGTCGGAGAGCTCATCTGAAATTTC 1755
QY 1570 GCAGATTTTAAAGCTCCTCAGATGTGAGAAGACCCATTCTCAACCAAGGGTGTTC 1629
Db 1756 GCCGACTTCATGACGAACCACTGACCGCTGAGAGTAAATCGGAAACAAAGAGGATCTTC 1815
QY 1630 ACAAGAGACAGACAAACCAACTCGTTGCTCATGTACTGAGA 1671
Db 1816 ACTCGCAGAGACAGCCCAAACTTCGGCCTTTATTTTGGCA 1857
```

```
RESULT 5
US-09-149-727-7
; Sequence 7, Application US/09149727
; Patent No. 6391547
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Kilian, Andrzej
; APPLICANT: Keese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/09/149,727
; EARLIER FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: US 60/058,263
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1887
; TYPE: DNA
; ORGANISM: Bacillus sp.
; US-09-149-727-7
```

```
Query Match 8.4%; Score 141.4; DB 3; Length 1887;
Best Local Similarity 45.4%; Pred. No. 2.7e-33;
Matches 734; Conservative 0; Mismatches 822; Indels 60; Gaps 4;
```

```
QY 120 CCAGATCTGTGCTACGAAGAAGACCCCTTCACTACAAACCACTTCTACGTTCCGAA 179
Db 208 CAAGAAATCCGCAACCATATCGGATATGTGTGTGACGAACGTGAGTTCAACGGTCCGC 267
QY 180 GNACTTTACAAAAACATCAGACTTTACTTGTGCTGCGTGAACACGACTGCGAGT 239
Db 268 CTATCTGAAGGATCAGCGTATCGTGTCCGCTTCGGCTCTGCAACTCACAAGCAATTGT 327
QY 240 CTTCTCAACGAGAGAAAGTGGAGAGAAATCAATTGAATACCTTCCCTTGAAGTAGA 299
Db 328 CTATGTCAATGTGAGCTGTGTGTGAGACAAAGGCGGATTCCTGCAATTGAAAGCGGA 387
QY 300 TGTGACGGGAAAGTGAATCCGGAGAGAACGAACTAAGGTTGTTGTGAGAACAGATT 359
```

Db 388 AATCAACAACCTGCTGCTGATGGCATGAATCGCGTCAACCGTCGCGCTGGACAACAATCCT 447
QY 360 GAAAGTGGGAGGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACAACCGTGGAATTTT 419
Db 448 CGACGATAGCACCCCTCCGGTGGGGCTGTACAGCGAGCGCCACGAAAGGGCTCGGAAA 507
QY 420 TGAAGTTTTCACCTGCAAACTTCGACTTCTCCCTACGGTGAATCATAGGCGTGT 479
Db 508 AGTCATTCTAACAAAGCCGAACCTTCGACTTCTTCAACTATGACAGGCTGCAACCGTCGGT 567
QY 480 TCTGATAGATTCAACAGACCAGCGAGGATACTCGACATCTGGGTGACACGAGTAGTC 539
Db 568 GAAAACTTACACGACCCCGTTTACGTACGTGAGGACATCTCGTTGTGACCGACTTCAA 627
QY 540 TGAACCGGAGAAACTTGAAGTGAAGTAGAAGTAGAGTCTCAGAAGAGCGGT 599
Db 628 TGGCCCAACCGGACTGTGACCTATACGGTGGACTTCAAGGCAAGCCGAGACCGTGAA 687
QY 600 GGGACAGAGATGACGATCAAACTTGGAGAGAGAGAAAGATTAGAACATCCAACAG 659
Db 688 AGTGTGGGTGCGTGAATGAGGAAGGCAAGTGTGCGAAGCACCGAGGGCTGAGCGGTAA 747
QY 660 ATTGTCGAAGGGAGTTTCATCTCGAAAAACGCCAGGTTCTGAGCCTCGAAGATCCATA 719
Db 748 CGTGAGATTTCGAATGTTCATCTCTGCGGAACCACTGAACACGTATCTTACCAGATCAA 807
QY 720 TCTTATCTCTCAAGGTGGAACCTTGAAGAAAGACGACTACCTCTGACATCGGAATCAG 779
Db 808 AGTGGAATCTGTGAACGACGAGCTGACCATCGATGTCTATGAAGAGCCGTTCCGGCTCG 867
QY 780 AACGATCAGCTGGGACGAGAGAGGCTCTATCTGAACGGGAAACCTGTTCTTTGAAGGG 839
Db 868 GACCGTGGAGTCAACGACGCGCAAGTTCTCTATCAACAACAAACCGTTCTACTTCAAAGG 927
QY 840 CTTTGAAGAACGACGAGGAATTCGCCGTTCTGGGCGAGGGCACTTTTATCCATTGATGAT 899
Db 928 CTTTGGCAACATGAGGACACTCTCTATCAACGCCGCTGGCTTTAAGCAAGCGACCAATGT 987
QY 900 AAAAGACTTCAACCTTCTGAAGTGAATCAACGCCAAATTCTTTCAGGACCTCTCACTATCC 959
Db 988 GATGATTTTCAATATCTCAATATGATCGGCGCCCAACAGCTTCCGACCGCACACTATCC 1047
QY 960 TTAAGTGAAGAGTGGCTGGATCTTGGCCGACAGACTCGGAATCCTTGTGATAGCAAGC 1019
Db 1048 GTACTCTGAAGAGTTGATGCGTCTTGGCGATCGCGAGGGTCTGGTGTGATCGACGAGAC 1107
QY 1020 CCCCACGTTGGTAT----- 1034
Db 1108 TCCGGCAGTTGGCGTGCACTCAACTTCATGCGCACCAAGGACTCGGCGAAGCGACCGA 1167
QY 1035 ---CACAAAGTACCACTACAATCCCGAGACTCAGAAAGTAGCAGAGACAACAATTAAGAG 1091
Db 1168 GCGCGTCAGTACTGGGAGAAAGATTCGGAAGTTGAGGCAACCATCAAGACGTTCTCCGTGA 1227
QY 1092 AATGATCGACAGACAAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAAGAAC 1151
Db 1228 ACTGGTGTCTGTGACAAAGAACCATCCAAAGCGTGTGATGTGAGCATCGCCAAAGAGGC 1287
QY 1152 AGAGTCCAAACCATTCAGAGCGCGAGGGTTTCTTCAAGCCCTTTATGAGACTGCGCAATGA 1211
Db 1288 GCGGACTGAGAGAGGGCGCGTACAGTACTTCAAGCCGTTGGTGAAGCTGACCAAGGA 1347
QY 1212 AATGGATCGA--ACAAGCCCGTTGTTCATGTGAGCATGATGAGCGCACCAAGAGAG 1268
Db 1348 ACTCGACCCACAGAGCGTCCGCTGACGATCGTGTGTTGTGATGAGCTAACCCCGAGAGAC 1407
QY 1269 AACAGAGAGCTGGCGCTGAAGTACTTGACATGCTGTGTGAACAGGTACTACGGCTG 1328
Db 1408 GGAACAAAGTCCCGAACTGA-----TTGACGTCAATCGCGTCAATCGCTATAACGGATG 1461
QY 1329 GTACATCTATCAGGAAAGATAGAGAG--GACTCAAGCTCTGGAAGAAAGACATAGA 1385

Db 1462 GTACTTCGATGGCGGTGATCTCGAAGCGGCCAAAGTCCATCTTCGCCAGGAATTTCAAGC 1521
QY 1386 AGAGCTCTATGCAAGGCGACAGAAAGCCCATCTTTGTACAGAAATTCCGTGCGGACGGAT 1445
Db 1522 GTGGAAACAAGCGTTGCCAGGAAAGCCGATCATGATCACTGATACGGCGACAGACCGT 1581
QY 1446 AGCTGGCATCCACTACGATCCACCTCAAAATGTTCTCGAAGAGTACCAAGCAGACTCGT 1505
Db 1582 TGGGGCTTTACGACATGATGATCAGTATGTTCAACGAGGAATATCAAGTGAAGTACTA 1641
QY 1506 TGAAGACGATCAGGCTCTTTTGAAGAAAGTACATCATTCGAAACACACGTTGGGC 1565
Db 1642 CAGGCGAACCAGCTGCTGTGATGATGATTGAGAACTTCGTGGGTGAGCAAGCGTGAA 1701
QY 1566 CTTTGCAGATTTTAAAGCTCTCAGAAATGTGAGAAAGCCCATTTCTCAACCAAGGGTGT 1625
Db 1702 CTTGCGGACTTCGCGACCTCTCAGGGCGGTGATGCGCGTCCAAGAAACAAAGAGGGCGT 1761
QY 1626 TTTCACAAGAGACAGACAACCCAACTCGTTGCTCATGTACTGAGAAAGACTGTGA 1681
Db 1762 GTTCACTGCTGACCGCAAGCCGAAGCTCGCGCGACAGTCTTTCGCGAGCGCTGA 1817

RESULT 6
US-09-270-957-27
; Sequence 27, Application US/09270957
; Patent No. 6641996
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL - GLUCURONIDASE GENES, GENE
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/09/270,957
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1888
; TYPE: DNA
; ORGANISM: Bacillus sp.
US-09-270-957-27

Query Match 7.7%; Score 130.4; DB 4; Length 1888;
Best Local Similarity 45.4%; Pred. No. 7.7e-30;
Matches 734; Conservative 0; Mismatches 822; Indels 61; Gaps 5;
QY 120 CCAGGATCTGTCCTAGGAAGAGACCTTTCACCTACAAACCACTTCTAGTTCGGA 179
Db 208 CAAGGAATCCGCAACCATATCGGATATGTCTGTACGAACGTGAGTTCAACGTCGCGC 267
QY 180 GNAACCTTCACAAAAACATCAGACTTCTTACTTGTCTGCGGTGAACACGAGCTCGAGGT 239
Db 268 CTATCTGAAGATCAGCGTATCTGCTCCGCTTCGGCTCTGCAACTCAAAAGCAATTGT 327
QY 240 CTTCTCAACGAGAGAAAGTGGAGAGATCAATTGAATACCTTCCCTTGAAGTAGA 299
Db 328 CTATGTCAATGTGAGCTGTGCTGAGGACCAAGGGCGGATTCCTGCAATTCGAAGCGGA 387
QY 300 TGTGACGGGGAAGTGAATCCGAGAGAGAACTCAGGGTGGTGTGAGAAACAGATT 359
Db 388 AATCAACAACCTCGCTGCGTGAATGCGGTCAACGCTCGCGTGAACAATCCT 447
QY 360 GAAAGTGGAGGATTTCCCTCGAAAGTTCCAGACAGCGGCACTCACACCGTGGGATTTT 419
Db 448 CGACGATAGCACCCCTCCGGTGGGGCTGTACAGCGGCGCAAGAGAGGGCGCTCGGAAA 507
QY 420 TGAAGTTTTCACCTGCAAACTTGACCTTCTCCCTACGGTGAATCATTAAGCGCTGT 479
Db 508 AGTCATTCTGAACAAGCCGAACCTTGACCTTCTTCAACTATGCAAGGCGCTGCAACCGTCCGT 567
QY 480 TCTGATAGAGTTCAACAGACACGCGAGGATACTCGACATCTGGGTGACACGAGTGAATC 539
Db 568 GAAATCTACACGACCCCGTTTACGTACGTGAGGACATCTCGGTTGTGACCGACTTCAA 627

QY	540	TGAACCGGAGAAAGAACTTGGAAAAGTGAAGATGAAGATGACTTCAGAGAAGCGGT	599
Db	628	TGGCCCAACCGGCACTGTGACCTATACGGTGCACTTCAAGGCCAAAGCCGAGACCGTGAA	687
QY	600	GCGACAGGAGATGACGATCAAACTTGGAGAGAGAGAAAAAGATTGAACATCCAAACAG	659
Db	688	AGTGTCCGTCGTGATGAGGAAGGCAAAAGTGTCGCAAGCACCGAGGGCCCTGAGCCGTAA	747
QY	660	ATTGCTCGAAGGGGAGTTCATCTTC -GAAAACGCCAAGTTCTGGAGACCTCGAAGATCCAT	718
Db	748	CGTGGAGATTCCGAAATGTCACTCTCTGGGAACCACTGAACACGTATCTTACCAGATCA	807
QY	719	ATCTTTATCCTCTCAAGGTGGAACCTTGAAAAAGACGAGTACACTCTGACATCGGAATCA	778
Db	808	AAGTGGAACTGGTGAACGACCGGACTGACCATGATGTCTATGAAGAGCCGTTCCGGCTGC	867
QY	779	GAACGATCAGCTGGGACGAGAGAAGGCTCTATCTGAACGGGAAAACCTGCTTTTGAAG	838
Db	868	GGAACCGTGGAAATCAACGACCGGCAAGTTCCTCATCAACAACAACCGTTTCTACTCAAG	927
QY	839	GCTTTGGAAAAGCAGAGGAATTCGCCGTTCTGGGGCAGGGCACCTTTATCCATTGATGA	898
Db	928	GCTTTGGCAAACATGAGGACACTCTTATCAACGCGCCGTGGCTTTAACGAAGCGAATG	987
QY	899	TAAAGACTTCAACCTTCTGAAGTGATCAACGCCGAATCTTTCAAGACCTCTCACTATC	958
Db	988	TGATGGATTCAATATCCTCAATGATCGGCGCCAACAGCTTCCGGACCGACACTATC	1047
QY	959	CTTACAGTGAAGGTGGCTGAACTCTTCCGACAGACTCGGAATCCTTGATAGACGAAG	1018
Db	1048	CGTACTCTGAAGAGTTGATGCGCTTGGCGGATGCGGAGGGTGTGCTGATCGACGAGA	1107
QY	1019	CCCCGCACGTTGGTAT-----	1034
Db	1108	CTCCGGCAGTTGGCGTGCACTCAACTTCATGSCCACCAGGGACTCGCGAAGGACGCG	1167
QY	1035	-----CACAAAGTACCACTACAATCCCGAGACTCAAGAAGATAGCAGAAGACAATAAGAA	1090
Db	1168	AGCGCGTCACTACCTGGAGAGAAGATTCCGACGTTTGAGCACCATCAAGACGTTCTCCGTG	1227
QY	1091	GAATGATGCACAGACACAAGAACCATCCAGTGTGATCATGTGGAAGTGTGGCGAAGCAAC	1150
Db	1228	AACGTGTGTCTCGTGAACAAGAACATCCAAAGCGTGTGATGTGGAAGCATCCCAACGAGG	1287
QY	1151	CAGAGTCCAAACCATCCAGAGCGGAGGGTTTCTTCAAGCCCTTATGAGACTGCCAATG	1210
Db	1288	CGCGACTGAGGAAGAGGGCGCGTACGAGTACTTCAAGCCGTTGGTGAAGTGAACCAAGG	1347
QY	1211	AAATGATCGA---ACACGCCCGTGTGCATGTGTGAGCATGTATGACGACACAGACGAGA	1267
Db	1348	AACTCGACCCACAGAAAGCGTCCGCTCACGATCGTGTGTGTGTGATGGCTAACCCCGAGA	1407
QY	1268	GAACAAGAGACGTGGCGCTGAAGTACTTTCGACATCGTCTGTGTGAACAGGTACTACGGCT	1327
Db	1408	CGGACAAAGTCGCCAACTGA-----TTGACGTCATCGCGCTCAATCGCTATAACGGAT	1461
QY	1328	GGTACATCTATCAGGGAAAGATAGAAGAG--GACTTCAAGCTCTGAAAAAGACATAG	1384
Db	1462	GGTACTTCGATGGCGGTGATCTGAAAGCGGCCAAAGTCCATCTCCGCCAGGAATTTACG	1521
QY	1385	AAGAGCTCTATGCAAGGCAAGAAAGCCCATCTTTGTCAAGAAATTCCGTCGGACGCGGA	1444
Db	1522	CGTGGAAACAAGCGTTGCCCAGGAAAGCCGATCATGATCATGTAGTACGGCGCAGACACCG	1581
QY	1445	TAGCTGGCATCCACTACGATCCAACTCAAAATGTTCTCCGAAGAGTACCAAGCAGAGCTCG	1504
Db	1582	TTGGGGGCTTTCAGGACATTGATTCAGTGATGTTTCAACCGAGGAATATCAAGTCGAGTACT	1641
QY	1505	TTGAAAAAGCAGATCAGGCTCCTTTGGAAAAAGACTACATCATCGGAACACACGCTGTGGG	1564
Db	1642	ACCAAGGCCGAACCAAGTCTGTTGATGAGTTTGAAGAACTTCTGGGTGAGCAAGCGTGGA	1701

Qy 1565 CCTTGCAGATTTTAAAGCTCTCAGAATGTGAGAAAGCCATTCTCAACCAACAAGGGTG 1624
 |||||
 Db 1702 ACTTCGGGACTTCGGGACTCTCAGGGGCGTGATGCCGCTCAAGGAAACAAGAGGGCG 1761
 |||||
 Qy 1625 TTTTCAAGAAGACAGACAAACCAACTCTGTGCTCATGTACTGAGAAAGACTGTGGA 1681
 |||||
 Db 1762 TGTTCACGTGTAACCGCAAGCTGAAGCTGCGCGGCACAGCTTTTTCGAGAGCCTGGA 1818
 |||||

RESULT 7

```

US-08-819-866-1/c
; Sequence 1, Application US/08819866
; Patent No. 5830698
;
; GENERAL INFORMATION:
;
; APPLICANT: REFF, Mitchell E.
;
; APPLICANT: BARNETT, Richard Spence
;
; APPLICANT: MCLACHLAN, Karen Retta
;
; TITLE OF INVENTION: NOVEL METHOD FOR INTEGRATING GENES AT
;
; TITLE OF INVENTION: SPECIFIC SITES IN MAMMALIAN CELLS VIA HOMOLOGOUS
;
; TITLE OF INVENTION: RECOMBINATION AND VECTORS FOR ACCOMPLISHING THE SAME
;
; NUMBER OF SEQUENCES: 2
;
; CORRESPONDENCE ADDRESSES:
;
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
;
; STREET: P.O. Box 1404
;
; CITY: Alexandria
;
; STATE: Virginia
;
; COUNTRY: United States
;
; ZIP: 22313-1404
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: Patentin Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/08/819, 866
;
; FILING DATE: 14-MAR-1997
;
; CLASSIFICATION: 435
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: Teskin, Robin L.
;
; REGISTRATION NUMBER: 35, 030
;
; REFERENCE/DOCKET NUMBER: 012712-352
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (703) 836-6620
;
; TELEFAX: (703) 836-2021
;
; INFORMATION FOR SEQ ID NO: 1:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 14683 base pairs
;
; TYPE: nucleic acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: DNA (genomic)
;
US-08-819-866-1

```

Query Match	Best Local Similarity	Score 112;	DB 2;	Length 14683;
Matches 472;	Conservative	0;	Mismatches 450;	Indels 60;
			Gaps 3;	
QY	757	TACACTCTGCATCGGAATCAGAACGATCAGCTGGACGAGAAAGAGCTTATCTGAAC	816	
Db	7348	TACCCGCTTCGCGTCGGCATCCGTCAGTGGCAGTGAAGGCGAACAGTTCTGATTAAAC	7289	
QY	817	GGGAAACCTGCTTTTGAAGGGCTTTGAAAGCAGAGAAATCCCCGTTCTGGGCAG	876	
Db	7288	CACAAACCGTTCTACTTTACTTTCGGCTTTGGTCGTCATGAAGATGCCGACTTGGCGCAAA	7229	
QY	877	GGCACCCTTTATCCATTGATGATTAAGAATTCACCTTCTGAAGTGATCAACGGCAAT	936	
Db	7228	GGATTGATTAACGTGCTGATGCTGACACGACCAAGCATTAATGAGCTGGATTAGGGCCAAC	7169	
QY	937	TCCTTCAGGACCTCTCAGTATCTTACAGTGAAGAGTGCGCTGGATCTTGCCGACAGACTC	996	
Db	7168	TCCTACCGTACTCGCATTACCTTTACGCTGAAGAGATGCTCGACTGGGCGAGATGAACAT	7109	


```

QY 997 GGAATCCTTGTGATAGACGAAGCCCC-----GCACGTTGGT 1032
Db 7108 GGCATCGTGGTGAATGATGAAACTGCTGCTGCGCTTTAACTCTCTTTAGGCATTGGT 7049
QY 1033 ATCACAAGGTACCACCTACTCAATCCGAGACTCAGAAGATAGCAGAAGACAAC----- 1083
Db 7048 TTCGAAGCGGGCAACAAGCCGAAAGAACTGTACAAAGAGAGGAGTCAACGGGAAACT 6989
QY 1084 -----ATAGAAGATGATCGACAGACACAGAACCATCCAGT 1122
Db 6988 CACCAAGCGCACTTACAGCGGATTAAGAGCTGATAGCGCGTGACAAAAACACCCCAAGC 6929
QY 1123 GTGATCATGTGAGTGTGCGGAACGAACGAGTCCAAACATCCAGACGCGGAGGGTTTC 1182
Db 6928 GTGTGATGTGAGTATTTGCCAACGAACCGGATACCCGTCGCAAGGTGACGGGAATAT 6869
QY 1183 TTCMAAGCCCTTTATGAGACTGCCAATGAATGATCGAACACGCCCGTTGTCAATGTG 1242
Db 6868 TTGCGGCCACTGGCGGAAGCAACGCGTAACTCGAACCGACGCGTCGATCACTCGCTC 6809
QY 1243 AGCATGATGAGCGCACCAAGAGAGAGAAAGAGACGTTGGCGCTGAAGTACTTCGACATC 1302
Db 6808 AATGTAATGTTCTGCGAGCGCTCACACCGATACCATCAGCGATCT-----CTTGAATGTG 6755
QY 1303 GTCTGTGTAACAGGTACTACGCGCTGTATCATCTATCAGGAAAGATAGAAGAGACTT 1362
Db 6754 CTGTGCTGAACCGTTATTAACGATGTATGTCMAAGCGCGCATTTGGAACCGCAGAG 6695
QY 1363 CAAGCTTGAAAAAGACATAGAAGAGCTCTATGCMAAGCACAGAAAGCCATCTTTGTC 1422
Db 6694 AAGTACTGGAAAAAGAACTCTGCGCTGGCAGAGAAACTGCATCAGCGGATTAATCATC 6635
QY 1423 ACAGAATTCGCTGCGAGCGCATAGCTGCGCATCCACTACGATCCACTCAATGTTCTTC 1482
Db 6634 ACCGAATACGGCGTGAATAGTACGTTAGCGCGGCTGCATCAATGTACACGACATGTGAGT 6575
QY 1483 GAAGAGTACCAAGCAGAGCTCGTTGAAGAACGATCAGGCTCCTTTGAAAAAAGACTAC 1542
Db 6574 GAAGAGTATCAGTGTGATGAGTGTGATGATCAACGCGCTCTTGAATGCGCTCAGCGCC 6515
QY 1543 ATCATCGGAACACACGTTGGGCGCTTTGAGATTTTAAAGTCTCCTCAGAATGTAGAGA 1602
Db 6514 GTGTCGCTGAACAGGTATGGAATTTGCGCGATTTTGCACCTCGCAGGACATATTGCGC 6455
QY 1603 CCCATTCTCAACCAAGGGTGTTCACAAGAGACAGACCAACCAACTCGTTGCTCAT 1662
Db 6454 GTTGGCGGTAAACAAGAAAGGATCTTCACTCGCGACCGCAAAACCGAAGTCCGCGCTTTT 6395
QY 1663 GTACTGAGAAGACTGTGAGTG 1684
Db 6394 CTGCTGCAAAAACGCTGAGACTG 6373

```

```

RESULT 8
US-09-023-715-1/c
; Sequence 1, Application US/09023715
; Patent No. 5998144
; GENERAL INFORMATION:
; APPLICANT: REFF, Mitchell E.
; APPLICANT: BARNETT, Richard Spence
; APPLICANT: MCLACHLAN, Karen Recta
; TITLE OF INVENTION: NOVEL METHOD FOR INTEGRATING GENES AT
; TITLE OF INVENTION: SPECIFIC SITES IN MAMMALIAN CELLS VIA HOMOLOGOUS
; TITLE OF INVENTION: RECOMBINATION AND VECTORS FOR ACCOMPLISHING THE SAME
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,715
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/819,866
; FILING DATE: 14-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 012712-352
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14683 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-023-715-1

```

```

Query Match 6.6%; Score 112; DB 2; Length 14683;
Best Local Similarity 48.1%; Pred. No. 1.6e-23;
Matches 472; Conservative 0; Mismatches 450; Indels 60; Gaps 3;

```

```

QY 757 TACACTCTGACATCGGAATCAGAACGATCAGCTGGACGAGAAAGGCTTATCTGAAC 816
Db 7348 TACCGGCTTCGCGTGGCATCCGCTCAGTGGCAGTGAAGGGCGAAGTTCTGATTTAAC 7289
QY 817 GGGAAACCTGCTTTTGAAGGGCTTTGAAAGCAGAGAAATCCCGTTCTGGGCGAG 876
Db 7288 CACAAACGTTCTTACTTACTGCTGTTGTGTCATGAAGATGCGACTTGCGGCAAA 7229
QY 877 GGCACCTTTATTCATGATGATGAATAAAGACTTCAACTTCTGAAGTGAATCAACGGCAAT 936
Db 7228 GGATTCGATTAACGTCGTGATGTGACGACCAACGATTAATGACTGATTAAGGCGCAAC 7169
QY 937 TCTTTCAAGACCTCTCACTATCTTACAGTGAAGAGTGCTGATCTTGCCGACAGACTC 996
Db 7168 TCTTACCGTACCTGCACTTACCCTTACGCTGAAGAGATGCTCGACTGGCAGATGAACAT 7109
QY 997 GGAATCCTTGTGATAGACGAAGCCCC-----GCACGTTGGT 1032
Db 7108 GGCATCGTGTGATGATGAATACTGTGCTGTGCGCTTTAACTCTCTTTAGGCATTGGT 7049
QY 1033 ATCACAAGGTACCACTACATCCGAGACTCAGAAGATAGCAGAAGACAAC----- 1083
Db 7048 TTCGAAGCGGGCAACAAGCCGAAAGAACTGTACAAAGAGAGGAGTCAACGGGAAACT 6989
QY 1084 -----ATAGAAGATGATCGACAGACACAGAACCATCCAGT 1122
Db 6988 CACCAAGCGCACTTACAGCGGATTAAGAGCTGATAGCGCGTGACAAAAACACCCCAAGC 6929
QY 1123 GTGATCATGTGAGTGTGCGGAACGAACGAGTCCAAACATCCAGACGCGGAGGGTTTC 1182
Db 6928 GTGTGATGTGAGTATTTGCCAACGAACCGGATACCCGTCGCAAGGTGACGGGAATAT 6869
QY 1183 TTCMAAGCCCTTTATGAGACTGCCAATGAATGATCGAACACGCCCGTTGTCAATGTG 1242
Db 6868 TTGCGGCCACTGGCGGAAGCAACGCGTAACTCGAACCCGACGCGCTCGATCACTCGCTC 6809
QY 1243 AGCATGATGAGCGCACCAAGAGAGAGAAAGAGACGTTGGCGCTGAAGTACTTCGACATC 1302
Db 6808 AATGTAATGTTCTGCGAGCGCTCACACCGATACCATCAAGCATCT-----CTTGAATGTG 6755
QY 1303 GTCTGTGTAACAGTACTACGCGCTGTGATCATCTATCAGGAAAGATAGAAGAGACTT 1362

```

```
Db      6754 CTGTGCTGAACCGTTATTACGGATGATATGTCCAAAGCGCGATTGGAAACGGCAGAG 6695
QY      1363 CAAGCTCTGGAAGAAAGACATAGAGAGCTCTATGCAAGGCACAGAAAGCCCATCTTTGTC 1422
Db      6694 AAGGTAAGTGAAGAAAGAACTCTGCGCTGGCAGAGAAACTGCATCAAGCCGATTATCATC 6635
QY      1423 ACAGAAATTCGGTCGCGACCGGATAGTGGCATCCACTACGATCCACTCAATGTCTCC 1482
Db      6634 ACCGAATACGCGGTGATACGTAGCCGGGTGCACCTCAATGTACACCGACATGTGAGT 6575
QY      1483 GAAGAGTACCAAGCAGAGCTCGTTGAAAAAGACGATCAGGCTCCTTTGAAAAAGACTAC 1542
Db      6574 GAAGAGTATCAGTGTGCATGGCTGATATGTATCACCGCGCTTTGATCGCGTCAGCGCC 6515
QY      1543 ATCATCGGAACACAGTGTGGCCTTTCAGATTTTAAGACTCCTCAGATGTGAGAA 1602
Db      6514 GTGCTCGGTGAACAGATATGGAATTCGCCGATTTGGCAGCCTCGCAAGGCATATTGCCG 6455
QY      1603 CCCATTCTCAACCAAGGGTGTTCACAAGAGACAGAACCCCAACTCGTTGCTCAT 1662
Db      6454 GTTGGCGGTAAACAAGAGGATCTTCACTCGCGACCGCAACCGAAGTGGCGGCTTTT 6395
QY      1663 GTACTGAGAAGACTGTGAGTG 1684
Db      6394 CTGCTGCAAAAAACGCTGAGCTG 6373
```

RESULT 9

```
US-09-343-485A-1/c
; Sequence 1, Application US/09343485A
; Patent No. 6413777
; GENERAL INFORMATION:
; APPLICANT: REEF, MITCHELL R.
; APPLICANT: BARNETT, RICHARD S.
; APPLICANT: MCILACHLAN, KAREN R.
; TITLE OF INVENTION: NOVEL METHOD FOR INTEGRATING GENES AT SPECIFIC SITES IN
; TITLE OF INVENTION: MAMMALIAN CELLS VIA HOMOLOGOUS RECOMBINATION AND
; FILE REFERENCE: 037003-0275807
; CURRENT APPLICATION NUMBER: US/09/343,485A
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/023,715
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: 08/819,866
; PRIOR FILING DATE: 1997-03-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 14683
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
; OTHER INFORMATION: referred to as "Desmond"
US-09-343-485A-1
```

```
Query Match      6.6%; Score 112; DB 3; Length 14683;
Best Local Similarity 48.1%; Pred. No. 1.6e-23;
Matches 472; Conservative 0; Mismatches 450; Indels 60; Gaps 3;
```

```
QY      757 TAACTCTGACATCGAATCAGAACGATCAGTCGGACGAGAAAGCGCTCTATCTGAAC 816
Db      7348 TACCGCTTCGCGTGGCATCCGTCAGTGAGGAAAGGCGAACAGTTCTGATTAAC 7289
QY      817 GGGAACTGTCTTTTGAAGGGCTTTGAAAGCAGAGAAATCCCGTTCTGGGGCAG 876
Db      7288 CACAAACCGTTCTACTTTACTGCTTGGTGTGTCATGAAGATGCGGACTTGCGGCAAA 7229
QY      877 GGCACTTTTATCATGTGATGATAAAGACTTCAACTTCTGAAGTGATCAACGCGAAT 936
Db      7228 GGATTCGATAACGTGCTGATGTGCAAGACCAAGCATTAATGAGCTGAGATTAGGCCAAC 7169
QY      937 TCTTTCAGGACCTCTCACTATCTTACAGTGAAGAGTGGCTGANTCTTGCCGACAGACTC 996
```

```
Db      7168 TCCTACCGTACCTCGCATTAACCTTAACGTGAAGAGATGTCGACTGGCGAGATGAACAT 7109
QY      997 GGAATCCTTGTGATAGACGAAGCCCC-----GCACGTTGCT 1032
Db      7108 GGCATCGTGTGATGATGATAAACTGCTGTCTGCTGCGCTTAACTCTCTTTAGGCATTGCT 7049
QY      1033 ATCACAAGGTACCACTACATCCCGAGACTCAGAAGATGACGAAGACAAC----- 1083
Db      7048 TTCGAGCGGGCAACAGCGGAAAGAACTGTACAAAGAGGCGAGTCMAAGGGGAAACT 6989
QY      1084 -----ATAAGAAATGATCGACAGACACAAGAACCATCCAGT 1122
Db      6988 CACCAAGCGCACTTACAGCGGATTAAGACTGATAGCGGTGACAAAAACCAACCAAGC 6929
QY      1123 GTGATCATGTGAGTGTGGGAAAGAACAGAGTCCAAACCATCCAGACGGGAGGTTTC 1182
Db      6928 GTGATGATGAGTATTTGCCAAGAAACCGATACCCGTCGCAAGGTGCACGGGAATAT 6869
QY      1183 TTCAAAGCCCTTTATGAGACTGCCAATGAATGATCGAACAAGCCCGTTGTCATGTG 1242
Db      6868 TTCGCGCCACTGGCGGAAGCAACGCTAACTGACCCGACGCGTCCGATCACTGCGTC 6809
QY      1243 AGCATGATGACGACCAAGACGAGAAAGAAAGACGAGTGGCGCTGAAGTACTTCGACATC 1302
Db      6808 AATGTAATGTTCTGCGACGCTCACACCGATACCATCAGCGATCT-----CTTTGATGTG 6755
QY      1303 GTCTGTGTAACAGTACTACGCGTGTGATCATCTATCAGGAAAGATAGAAGAGACTT 1362
Db      6754 CTGTGCTGAACCGTTATTACGATGTATGTCCAAAGCGCGATTGGAAACGGCAGAG 6695
QY      1363 CAAGCTCTGGAAGAAAGACATAGAAAGCTCTATGCAAGGCAAGAAAGCCCATCTTTGTC 1422
Db      6694 AAGTACTGGAAGAAAGAACTTCTGCGCTGGCAGAGAAAGTGCATCAGCCGATTATCATC 6635
QY      1423 ACAGAAATTCGGTCGCGACCGGATAGTGGCATCCACTACGATCCACTCAATGTCTCC 1482
Db      6634 ACCGAATACGCGGTGATAGCTTAGCCGGGTGCACCTCAATGTACACCGACATGTGAGT 6575
QY      1483 GAAGAGTACCAAGCAGAGCTCGTTGAAAAAGACGATCAGGCTCCTTTGAAAAAGACTAC 1542
Db      6574 GAAGAGTATCAGTGTGCATGGCTGATATGTATCACCGCGCTTTGATCGCGTCAAGGCC 6515
QY      1543 ATCATCGGAACACAGTGTGGGCTTTGAGATTTTAAGACTCCTCAGAAATGTGAGAA 1602
Db      6514 GTGCTCGGTGAACAGTATGGAATTTGCGGATTTTGCACTCGCAAGGCATATTGCCG 6455
QY      1603 CCCATTCTCAACCAAGGGGTTCACAAGAGACAGAACCAACCAACTCGTTGCTCAT 1662
Db      6454 GTTGGCGGTAAACAAGAGGATCTTCACTGCGGACCGCAACCGAAGTGGCGGCTTTT 6395
QY      1663 GTACTGAGAAGACTGTGAGTG 1684
Db      6394 CTGCTGCAAAAAACGCTGAGCTG 6373
```

```
RESULT 10
US-08-723-624-18
; Sequence 18, Application US/08723624
```

```
; Patent No. 5861277
; GENERAL INFORMATION:
; APPLICANT: Rose, Alan B.
; APPLICANT: Last, Robert L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING
; TITLE OF INVENTION: THE EXPRESSION OF GENES IN PLANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSER: ARNOLD, WHITE & DURKEE
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/723,624
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: BTIP:002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3035 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-723-624-18

```

```

Query Match      6.6%; Score 110.8; DB 2; Length 3035;
Best Local Similarity 47.4%; Pred. No. 1.4e-23;
Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

```

```

QY 757 TACACTCTGACATCGAATCAGAAGCATCAGCTGGACGAGAAGAGCTCTATCTGAAC 816
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2016 TACCCGCTTCGCGTCGGCATCCGTCAGTGGCAGTGAAGGGCGAAGCTTCCTGATTAAC 2075

QY 817 GGGAAACCTGTCTTTTGAAGGCTTGAAGACGAGGAATTCCCGTTCTGGGCGAG 876
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2076 CACAACCGTTCTACTTACTGCTTGCTGCTGATGAAGATGCGACTTGCGTGCAAA 2135

QY 877 GGCACCTTTTATCCATGATGATAAAGACTTCAACTTCTGAAGTGATCAAGCGAAT 936
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2136 GGATTCGATAACGTGCTGATGTGACAGCACGACATTAATGACTGATTGGGCCAAC 2195

QY 937 TCTTTCAGGACCTCTCATCTATCTTACAGTGAAGATGGCTGATCTTCCGACAGACTC 996
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2196 TCCTACCGTACCTCGCATTAACCTTAGCTGAAGAGATGCTGACTGGGCAGATGAACAT 2255

QY 997 GGAATCCTTGTGATAGCAAGCCCGCAGCTTGATCACAAGT----- 1042
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2256 GGCATCGTGTGATGATGAAACTGCTGCTGCGCTTAACTCTCTTTAGGCAATTGCT 2315

QY 1043 -----ACCACTACAATCCCGAGACT 1062
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2316 TTGGAAGCGGGCAACAAGCCGAAGACTGTACAGCGAAGAGGCAAGCGGGAACACT 2375

QY 1063 CAGAAGATAGCAGAAGACACATAGAAGATGATCGACAGACACAGAACCATCCCACT 1122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2376 CAGCAAGCGCACTTACAGCGCATTAAGAAGCTGATAGCGCGTGACAAAAACCAACCAAGC 2435

QY 1123 GTGATCATGTGAGTGTGGCGAAGCAACAGAGTCCAAACCATCCAGACGGGAGGGTTTC 1182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2436 GTGGTATGTGAGTATTGCCAACGACCGGATACCCGTCGCAAGGTGCACGGGAATAT 2495

QY 1183 TTCAAAGCCCTTTATGACTGCCAATGAATGATGGAACACGCCCCGTTGTCAATGCTG 1242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2496 TTCGCCCACTGGCGGAAGCAACGCGTAACCTGACCCGACGCGTCGATCACTCGGCTC 2555

QY 1243 AGCATGATGAGCAGCAGACGAGAGAAACAAGAGAGCTGGCGCTGAAGTACTTCGACATC 1302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2556 AATGTAATGTTCTGCGACGCTCACACCGATACCATCAGCGATCT-----CTTGATGTG 2609

QY 1303 GTCTGTGTGAACAGGTAATAAGGCTGTATCATCTATCAGGGGAAGATAGAAGAGACTT 1362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2610 CTGTGCTGAACCGTTATTACGGATGTATGTCCAAAGCGGCGGATTTGAAACGGCAGAG 2669

QY 1363 CAAGTCTTGAAAAAGACATAGAAGAGCTTATGCAAGGCACAGAAAGCCCATCTTTGTC 1422

```

```

Db 2670 AAGGTAAGTGAAGAAAGAACTTCTGGCCTTGACAGAGAAAGTGCATCAGCCGATTAATCATC 2729
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1423 ACAGATTGCGTGGGACGCGATAGCTGGCATCCATCAGATCCACCTCAATGTCTCC 1482
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2730 ACCGATAACGCGGTGATACGTTAGCCGGCTGCACTCAATGTACCCGACATGTGAGT 2789

QY 1483 GAAGAGTACCAAGCAGAGCTCGTTGAAAAGACGATCAGGCTCCTTTGAAAAAAGACTAC 1542
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2790 GAAGAGTATCAGTGTGCATGGCTGATATGATATCACCAGCTCTTGATCGCGTCAGCGCC 2849

QY 1543 ATCATCGGAACACACGTGTGGGCTTTGACAGATTTTAAGACTCCTCAGATGTGAGAAGA 1602
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2850 GTCGTCGGTGAACAGATATGGAATTTCCCGATTTTGCAGACCTCGCAAGGCATATTGCCG 2909

QY 1603 CCCATTCTCAACCAAGGCTGTTTCAAGAGACAGACACCAACCAACTCGTGTCTCAT 1662
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2910 GTTGGCGGTAAACAGAAAGGATCTTCACTCGCGACCGCAACCGAAAGTCGGCGCTTTT 2969

QY 1663 GTACTGAGAGAGACTGTGAGTG 1684
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2970 CTGCTGCAAAAACGCTGGAAGT 2991

```

```

RESULT 11
US-08-630-820-5
; Sequence 5, Application US/08630820
; Patent No. 6008023
; GENERAL INFORMATION:
; APPLICANT: OPPER, Martin
; APPLICANT: BOSSLET, Klaus
; APPLICANT: CZECH, Joerg
; TITLE OF INVENTION: CYTOPLASMIC EXPRESSION OF ANTIBODIES,
; TITLE OF INVENTION: ANTIBODY FRAGMENTS AND ANTIBODY FRAGMENT FUSION MOLECULES
; TITLE OF INVENTION: IN E. COLI
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,820
; FILING DATE: 10-APR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 19513676.4
; FILING DATE: 11-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: GRANADOS, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 18748/306
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3169 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterobacteriaceae: Escherichia coli

```

STRAIN: PRAJ210
IMMEDIATE SOURCE:
CLONE: pTrc99 dicistr. Fab/E.c.-Beta-Gluc
FEATURE:
NAME/KEY: CDS
LOCATION: 3..641
FEATURE:
NAME/KEY: CDS
LOCATION: 666..3162
US-08-630-820-5

Query Match 6.6%; Score 110.8; DB 3; Length 3169;
Best Local Similarity 47.4%; Pred. No. 1.5e-23;
Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

QY 757 TAACTCTGGACATCGAATCAGAACGATCAGCTGGGACGAGAAAGGCTCTATCTGAAC 816
DB 2145 TACCCGCTTCGCGCGCATCCGGTCAGTGGCAGTGAAGGGCGAACAATTCTGATTAAAC 2204
QY 817 GGGAAACCTGCTTTTGAAGGGCTTTGAAAGCAGAGGAATTCCTTCTGGGGCAG 876
DB 2205 CACAAACCGTTCTACTTTACTGCTTGGTTCGTCAATGAAGATCGGACTTACGTGCCAAA 2264
QY 877 GGCACCTTTTATCATGTATGATATAAAGCTTCAACCTTCTGAAGTGATCAACCGCAAT 936
DB 2265 GGATTCGATTAACGTCGTGATGTGCAAGACACGCAATTAATGATGATGGGCCAAC 2324
QY 937 TCTTTCAGGACCTCTCACTATCTTACAGTGAAGATGGCTGATCTTGGCAGACTC 996
DB 2325 TCCTAACGTAACCTCGCATTTACCTTACGCTGAAGAGATGCTGACTGGGAGATGAACAT 2384
QY 997 GGAATCCTTGTGATAGACGAAGCCCGCAGCTTGTATCAAAAGT----- 1042
DB 2385 GGATCGTGTGATGATGAAACTGCTGCTGCTTTAACTCTTTAGGCAATGGT 2444
QY 1043 -----ACCACTACAATCCCGAGACT 1062
DB 2445 TTGGAACGGGCAACAGCCGAAAGAACTGTACAGCGAAGAGCAGTCAACGGGGAACCT 2504
QY 1063 CAGAGATAGCAGAAACAACATAAAGAATGATGACAGACACAAGAACCATCCAGT 1122
DB 2505 CAGCAAGCGCACTTAAGCGCATTAAGAAGCTGATAGCGCGTGACAAACCAACCAAGC 2564
QY 1123 GTGATCATGTGAGTGTGGCAAGCAACAGAGTCCCAACCATCCAGACGCGAGGTTTC 1182
DB 2565 GTGTGATGTGAGTATTGCCAAGCAACCGATACCCGTCGCAAGGTGACGGAATAT 2624
QY 1183 TTCAAAGCCCTTTATGAGACTGCAATGAATGATCGAACACGCCCGTTGTGATGTG 1242
DB 2625 TTGCGCGCACTGGCGGAAGCAACGCGTAACCTGACCCGACGCGTCCGATCACCTGCGTC 2684
QY 1243 AGCATGATGAGCAGCAGACGAGAGAAACAAGAGAGTGGCGCTGAAGTACTTGACATC 1302
DB 2685 AATGTAATGTTCTGCGACGCTCACACCGATACCATCAGCGATCT-----CTTGATGTG 2738
QY 1303 GTCTGTGTAACAGGTACTACGCTGTGATCATCTATCAGGAAGATAGAAGAGACTT 1362
DB 2739 CTGTGCTGAACCGTTATTACGAGATGATGTCCAAAGCGGCGATTGGAAACGGGAGAG 2798
QY 1363 CAAGCTCTGAAAAAGACATAGAAGAGCTCTATGCAAGGACACAGAAAGCCATTTGTC 1422
DB 2799 AAGGTAAGTGAAGAAAGAACTTCTGCGCTGGCAGAGAAAGTGCATCAGCCGATTAATC 2858
QY 1423 ACAGAAATTCGTCGGAACGCGATAGCTGGCATCCATCAGATCCACTCAATGTTCTCC 1482
DB 2859 ACCGAATACGCGGTGATACGTTAGCGGGGCTGCACTCAATGTAACCGACATGTGAGT 2918
QY 1483 GAAGAGTACCAAGCAGAGCTCGTTGAAGAGCAGATCAGGCTCCTTTGAAAAAGACTAC 1542
DB 2919 GAAGAGTATCAGTGTGATGCTGATGATATGATCAACCGCGCTTTGATCGCGTACGCC 2978
QY 1543 ATCATCGGAACACAGCTGTGGCGCTTTCAGATTTTAAGACTCTTCAGATGTGAAGA 1602

DB 2979 GTGTCGGTGAACAGGTATGGAATTTCCCGCATTTTGGACCTCGCAAGGCATATTGCCG 3038
QY 1603 CCCATTCTCAACCAAGAGGTGTTTTCACAAGACAGACACAACCAACTCGTGTCTCAT 1662
DB 3039 GTTGGCGGTAAAGAAAGGATCTTCACTGCGGACCGCAAAACGAAGTCGGCGGCTTT 3098
QY 1663 GTACTGAGAGACTGTGAGTG 1684
DB 3099 CTGCTGCAAAAACGCTGACTG 3120

RESULT 12
US-09-273-453-5
Sequence 5, Application us/09273453
Patent No. 6602688

GENERAL INFORMATION:
APPLICANT: OPPER, Martin
BOSSLET, Klaus
CZECH, Joerg
TITLE OF INVENTION: CYTOPLASMIC EXPRESSION OF ANTIBODIES,
ANTIBODY FRAGMENTS AND ANTIBODY FRAGMENT FUSION MOLECULES
IN E. COLI

NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/273,453
FILING DATE: 22-Mar-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,820
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: GRANADOS, Patricia D.
REGISTRATION NUMBER: 33,683
REFERENCE/DOCKET NUMBER: 18748/306
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3169 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterobacteriaceae: Escherichia coli
STRAIN: PRAJ210
IMMEDIATE SOURCE:
CLONE: pTrc99 dicistr. Fab/E.c.-Beta-Gluc
FEATURE:
NAME/KEY: CDS
LOCATION: 666..3162
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-273-453-5
Query Match 6.6%; Score 110.8; DB 4; Length 3169;

Best Local Similarity 47.4%; Pred. No. 1.5e-23;
 Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

QY	757	TACACTCTGACATCGGAATCAGAACGATCAGCTGGAGCAGAAAGGCTCTATCTGAAC	816
DB	2145	TACCGCTTCGCGTCGGCATCCGCTAGTGGCAGTGAAGGCGCAACAGTTCCTGATTAA	2204
QY	817	GGGAACCTGCTCTTTTGAAGGCTTTGAAAGCAGAGAAATCCCGTTCTGGGCGAG	876
DB	2205	CACAAACCTTCTACTTACTGCTTGTGCTGCTATGAAGATCGGACTTAAGTGGCAAA	2264
QY	877	GGCACCCTTTATTCATGATGATAAAAGCTCAACCTTCTGAAGTGAATCAACGCAAT	936
DB	2265	GGATTGATTAACGTGCTGATGCTGACAGCAGCAGATTAATGACTGATGGGCGCAAC	2324
QY	937	TCTTTCAGACCTCTCACTATCCTTACAGTGAAGAGTGGCTGATCTTGGCAGACATC	996
DB	2325	TCCTACCGTACCTCGCATTAACCTTACGCTGAAGAGATGCTCGACTGGCAGATGAACAT	2384
QY	997	GGAATCCTTGTGATAGACGAAGCCCGCAGCTTGTATCAAGT-----	1042
DB	2385	GGCATCGTGTGATGATGAATACTGCTGCTGCTTAACTCTCTTAAAGCATTTGT	2444
QY	1043	-----ACCATACTACCTCCGAGACT	1062
DB	2445	TTGGAAGCGGGCAACAAGCCGAAGAAGTGTACAGCGAAGAGGAGTCAACGGGAAACT	2504
QY	1063	CAGAAGATAGCAGAAGACATAGAAGATGATCGACAGACACAAAGCATCCCACT	1122
DB	2505	CAGCAAGCGCACTTACAGCGGATTAAGAAGCTGATAGCCGCTGACAAAACCAACCGAAGC	2564
QY	1123	GTGATCATGTGAGTGTGGCGAAGCAGACAGAGTCCCAACCATCCAGACGCGGAGGTTTC	1182
DB	2565	GTGATGATGTGAGTGTGGCGAAGCAGACAGAGTCCCAACCATCCAGACGCGGAGGTTTC	1182
QY	1183	TTCAAGCCCTTATGAGACTGCGCAATGAATGATCGAACAAGCCCGTTGTATGTTG	1242
DB	2625	TTCCGCGCACTGGCGGAGCAAGCGCTAACTCGACCCGAGCGCTCCGATCACTCCGCTC	2684
QY	1243	AGCATGATGAGCGCACCGAGAGAGACAGAGAGCGTGGCGCTGAAGTACTTCGACATC	1302
DB	2685	AATGTAATGTTCTGCGAGCTCACAACCGATACCATCAGCGATCT-----CTTGATGTG	2738
QY	1303	GCTGTGTGAACAGGTAATCGGCTGTATCATCTATCAGGGAAGATGAAGAAAGACTT	1362
DB	2739	CTGTGCTGAACCGTTATTAAGGATGTATGTCGAAAGCGCGGATTGGAAACGCGCAGAG	2798
QY	1363	CAGCTCTGGAAGAAAGACATAGAAGACTCTATGCAAGGCAAGAAAGCCCATCTTTGTC	1422
DB	2799	AAGTAATGGAAGAAAGAACTTCTGCGCTGGCAGAGAACTGCATCAGCCGATTATCATC	2858
QY	1423	ACAGAAATTCGGTGGCGGAGTAGCTGCGATCCACTACGATCCACTCAATGTTCCTCC	1482
DB	2859	ACCGAATACGGCGTGGATGATGCTGAGTGTATCAACCGCTCTTGAATCGCTCAGCGCC	2918
QY	1483	GAAAGATCAAGCAGAGCTGTTGAAGAGAGATCAGGCTCTTTGAAAAAGAGACTAC	1542
DB	2919	GAAAGATCAAGTGTGATGCTGAGTGTATCAACCGCTCTTGAATCGCTCAGCGCC	2978
QY	1543	ATCATCGGAACACACGCTGGGCTTTGACAGATTTTAAGACTCTCAGATGTGAAGA	1602
DB	2979	GTCGCTGTGAACAGATGAATTTGCGCGATTTTGGCAGCTCGAAGCATATTGCGC	3038
QY	1603	CCCATCTCAACACAGAGGCTTTTCAAGAGAGAGACAAACCAACTCGTTGCTCAT	1662
DB	3039	GTTGGCGGTAAAGAAAGGATCTTCACTCGCAGCGCAAAACGAAAGTGGCGGCTTTT	3098
QY	1663	GTACTGAGAAGACTGTGAGTG 1684	
DB	3099	CTGCTGCAAAAACGCTGACTG 3120	

RESULT 13

US-08-723-624-19
 ; Sequence 19, Application US/08723624
 ; Patent No. 5861277
 ; GENERAL INFORMATION:

; APPLICANT: Rose, Alan B.
 ; APPLICANT: Laet, Robert L.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING
 ; TITLE OF INVENTION: THE EXPRESSION OF GENES IN PLANTS
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: ARNOLD, WHITE & DURKEE
 ; STREET: P. O. Box 4433
 ; CITY: Houston
 ; STATE: TX
 ; COUNTRY: USA
 ; ZIP: 77210

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/723, 624
 ; FILING DATE: Concurrently Herewith
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:

; NAME: Parker, David L.
 ; REGISTRATION NUMBER: 32,165
 ; REFERENCE/DOCKET NUMBER: BTIP:002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 512/418-3000
 ; TELEFAX: 512/474-7577

; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3824 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-723-624-19

Query Match 6.6%; Score 110.8; DB 2; Length 3824;

Best Local Similarity 47.4%; Pred. No. 1.7e-23;

Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

QY	757	TACACTCTGACATCGGAATCAGAACGATCAGCTGGAGCAGAAAGGCTCTATCTGAAC	816
DB	2805	TACCGCTTCGCGTGGCATCCGCTCAGTGGCAGTGAAGGCGCAACAGTTCCTGATTAA	2864
QY	817	GGGAACCTGCTCTTTTGAAGGCTTTGAAAGCAGAGAAATCCCGTTCTGGGCGAG	876
DB	2865	CACAAACCGTTCTACTTACTGCTTTGCTGCTGATGAAGATGCGGACTTGGGCGCAAA	2924
QY	877	GGCACCCTTTATTCATGATGATTAAGAACTTCAACCTTCTGAAGTGAATCAACGCAAT	936
DB	2925	GGATTGATAACGCTGCTGATGCTGACAGCAGCAGCATTAATGACTGATTGGGCGCAAC	2984
QY	937	TCTTTCAGGACCTCTCACTATCTTACAGTGAAGAGTGGCTGATCTTGGCAGACACTC	996
DB	2985	TCCTAACGTAACCTCGCATTAACCTTACGCTGAAGAGATGCTCGACTGGGCGAGTGAACAT	3044
QY	997	GGAATCCTTGTGATAGACGAAGCCCGCAGCTTGTATCAAGT-----	1042
DB	3045	GGCATCGTGTGATGATGAATACTGCTGCTGCTTAACTCTCTTAAAGCATTTGT	3104
QY	1043	-----ACCACTAATCCGAGACT	1062
DB	3105	TTGGAAGCGGGCAACAAGCCGAAGAAGTGTACAGCGAAGAGGAGTCAACGGGAAACT	3164
QY	1063	CAGAAGATAGCAGAAGACATAGAAGATGATCGACAGACACAAAGCATCCCACT	1122
DB	3165	CAGCAAGCGCACTTAACAGCGGATTAAGAAGCTGATAGCGCGTGAACAAAACCAACCGAAGC	3224
QY	1123	GTGATCATGTGAGTGTGGCGAAGCAGAACGATCCAAACCATCCAGACGCGGAGGTTTC	1182

Db 3225 GTGTGATGTGAGTAATTGCCAACGAACCGGATACCCGTCCGCAAGGTGCACGGGAATAT 3284
QY 1183 TTCAAAGCCCTTTATGAGACTGCCAATGAATGATCGAACACGCCCGTTGTTCATGTG 1242
Db 3285 TTCCGCCCACTGGCGGAAGCAACGCCGTAAACTCGACCCGACGCGATTCGATCACCCTGCGTC 3344
QY 1243 AGCATGATGACGACCAAGACGAGAGAAACAAGAGAGTGGCGCTGAAGTACTTCGACATC 1302
Db 3345 AATGTAATGTTCTGCGACGCTCACACCGATACCATCAGCATCT-----CTTGAATGTG 3398
QY 1303 GTCTGTGAACAGGTACTACGCGCTGTACATCTATCAGGAAGATAGAAGAAAGACTT 1362
Db 3399 CTGTGCTGAACCGTTATTAACGATGATGTATGTCAAAGCGCGAATTGGAAACGGCAGAG 3458
QY 1363 CAAGCTCTGAAAAAGACATAGAAAGCTCTATGCAAGGCACAGAAACCCATCTTTGTC 1422
Db 3459 AAGGTACTGAAAAAAGAACTTCTGGCCTGGCAGAGAAACTGCATCAGCCGATTAATCATC 3518
QY 1423 ACAGATTTCGTCGCGACGCGATAGCTGGCATCCACTCACTCAATGTCTCC 1482
Db 3519 ACCGAATACGCGGTGATACGTTAGCGGGGTGCACTCAATGTACACCGACATGTGAGT 3578
QY 1483 GAAGAGTACCAAGCAGAGCTGTTGAAAAAGACGATCAGGCTCCTTTGAAAAAAGACTAC 1542
Db 3579 GAAGAGTATCAGTGTGATGCGTGATGTATCACCAGCGCTTTGATCGCGTACGCGCC 3638
QY 1543 ATCATCGGAACACAGCTGTGGCTTTGAGATTTTAAGACTCTTCAGATGTGAAGA 1602
Db 3639 GTCTGCGGTGAACAGGATGAATTTCCGCGATTTTCCGACCTCGCAAGGCATATTTGCCG 3698
QY 1603 CCCATTCTCAACCAAGGGGTGTTTCAAGAGACAGACAAACCAACTCGTTGCTCAT 1662
Db 3699 GTTGGCGGTAAACAAGAAAGGATCTTCACTCGCAGCCGCAAAACCGAAGTCCGCGCTTTT 3758
QY 1663 GTACTGAGAAGACTGTGAGTG 1684
Db 3759 CTGCTGCAAAAACGCTGGAAGTG 3780

RESULT 14

US-09-893-525-36
; Sequence 36, Application US/09893525
; Patent No. 6753167
; GENERAL INFORMATION:
; APPLICANT: Moloney, Maurice M.
; TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies
; FILE REFERENCE: 9369-172
; CURRENT APPLICATION NUMBER: US/09/893, 525
; PENDING FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 09/210, 843
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: US 08/846, 021
; PRIOR FILING DATE: 1997-04-25
; PRIOR APPLICATION NUMBER: US 08/366, 783
; PRIOR FILING DATE: 1994-12-30
; PRIOR APPLICATION NUMBER: US 08/142, 418
; PRIOR FILING DATE: 1993-11-16
; PRIOR APPLICATION NUMBER: US 07/659, 835
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36
; LENGTH: 4652
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Phas-GUS-phas
; NAME/KEY: CDS
; LOCATION: (1548)..(3359)
; OTHER INFORMATION:
US-09-893-525-36

Query Match 6.6%; Score 110.8; DB 4; Length 4652;
Best Local Similarity 47.4%; Pred. No. 1.9e-23;
Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;
QY 757 TACACTGTGACATCGAATCAGAAGCATGAGCTGGACGAGAAAGAGCTCTATCTGAAC 816
Db 2343 TACCCTGTCGCTGGCAATCCCGTCAGTGGCAGTGAAGGCCAACAGTTCCTGATTAAC 2402
QY 817 GGGAAACCTGTCTTTTGAAGGGCTTTGAAAGCAGAGAAATCCCCGTTGCGGCGAG 876
Db 2403 CACAAACCGTTCTACTTACTGTGCTTTGGTGTGATGAAGATGCGACTTACGTGGCAAA 2462
QY 877 GGCACCTTTTATCCATTGATGATAAAGACTTCAACCTTCTGAAGTGAATCAACCGAAT 936
Db 2463 GGAATTGATTAAGTGTGATGATGACAGCAGCATTAAATGAACTGATTTGGGCCAAC 2522
QY 937 TCTTTACGACCTCTCACTATCCTTACAGTGAAGAGTGGCTGGATCTTGCGACAGACTC 996
Db 2523 TCCTACCGTACTCGCATTAACCTTACGCTGAAGAGATGCTCGAAGTGGCAGATGAACAT 2582
QY 997 GGAATCCTTGTGATAGACGAAGCCCGCAGCTTGGTATCACAAAGT----- 1042
Db 2583 GGCATCGTGTGATTTGATGAAACTGCTGTGCTTTCGCTCTCTTAAAGCATTTGT 2642
QY 1043 -----ACCACTACAATCCGAGACT 1062
Db 2643 TTCGAAGCGGGCAACAAGCCGAAGAAGCTGTACAGCGAAGAGGAGTCAACGGGAAACT 2702
QY 1063 CAGAAGATAGCAGAAGACAAACATAAGAAAGATGATCGACAGACACAGAAGCAATCCAGT 1122
Db 2703 CAGCAAGCGCACTTACAGCGATTAAGAGCTGATAGCGCGTGAACAAAAACCAACCAAGC 2762
QY 1123 GTGATCATGTGAGTGTGGCGAAGCAAGCAGAGTCCCAACATCCAGACGCGGAGGTTTC 1182
Db 2763 GTGTGATGTGAGTATTTGCCAAGCAACCGGATACCGGTCCGCAAGGTGCACGGGAATAT 2822
QY 1183 TTCAAAGCCCTTTATGAGACTGCCAATGAATGATCGAACAACGCCCGCTGTTCATGTG 1242
Db 2823 TTCCGCCACTGGCGGAAGCAACGGTAACCTCGAACCCGCTCCGATCACTGCGTC 2882
QY 1243 AGCATGATGACGCAACCAAGCAGAGAAACAAGAGACGTGGCCTGGAAGTACTTCGACATC 1302
Db 2883 AATGTAATGTTCTGCGACGCTCACACCGATACCATCAGCGATCT-----CTTTGATGTG 2936
QY 1303 GTCTGTGTAACAGGTACTACGGCTGTGTACATCTATCAGGGAAGGATAGAAGAAAGACTT 1362
Db 2937 CTGTGCTGAACCGTTATTAACGATGTGTATCCAAAGCGCGATTTGGAACGGCAGAG 2996
QY 1363 CAAGCTCTGAAAAAGACATAGAAAGAGCTCTATGCAAGGCACAGAAAGCCCATCTTTGTC 1422
Db 2997 AAGGTACTGAAAAAAGAACTTCTGCGCTGGCAGGAGAAACTGCATCAGCCGATTAATC 3056
QY 1423 ACAGAAATTCGTCGCGGACCGCATAGCTGGCATCCACTACGATCCACCTCAATGTCTCC 1482
Db 3057 ACCGAATACGCGGTGATACGTTAGCCGGCTGCACCTCAATGTACACCGCATGTGAGT 3116
QY 1483 GAAGAGTACCAAGCAGAGCTGTTGAAAAAGCAGTCAAGCTCTTTTGAAAAAAGACTAC 1542
Db 3117 GAAGATATCAGTGTGATGCTGATATGTATCACCAGCTCTTTGATCGCGTCAGCGCC 3176
QY 1543 ATCATCGGAACACAGCTGTGGGCTTTGCAATTTTAAGACTCTCAGAAATGTGAAGA 1602
Db 3177 GTGCTGGTGAACAGGTATGGAATTTGCGGATTTTGGACCTCGCAAGGCATATTGCGC 3236
QY 1603 CCCATTCTCAACCAAGGCTGTTTTCACAAGAGACAGACAAACCAACTCGTTGCTCAT 1662
Db 3237 GTTGGCGGTAAACAAGAAAGGATCTTCACTCGCAGACCGCAAAACGAAAGTGGCGGCTTTT 3296
QY 1663 GTACTGAGAAGACTGTGAGTG 1684
Db 3297 CTGCTGCAAAAACGCTGGAAGTG 3318

```

RESULT 15
US-09-893-525-41
; Sequence 41, Application US/09893525
; Patent No. 6753167
; GENERAL INFORMATION:
; APPLICANT: Moloney, Maurice M.
; APPLICANT: Van Rooijen, Gijb
; TITLE OF INVENTION: Preparation of Heterologous Proteins on Oil Bodies
; FILE REFERENCE: 9369-172
; CURRENT APPLICATION NUMBER: US/09/893,525
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 09/210,843
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: US 08/846,021
; PRIOR FILING DATE: 1997-04-25
; PRIOR APPLICATION NUMBER: US 08/366,783
; PRIOR FILING DATE: 1994-12-30
; PRIOR APPLICATION NUMBER: US 08/142,418
; PRIOR FILING DATE: 1993-11-16
; PRIOR APPLICATION NUMBER: US 07/659,835
; PRIOR FILING DATE: 1991-02-22
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
; LENGTH: 5390
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: phas-caleo-GUS-phas
; NAME/KEY: CDS
; LOCATION: (1548)..(4097)
; OTHER INFORMATION:
;
US-09-893-525-41

```

Query Match 6.6%; Score 110.8; DB 4; Length 5390;
 Best Local Similarity 47.4%; Pred. No. 2.1e-23;
 Matches 465; Conservative 0; Mismatches 457; Indels 60; Gaps 2;

```

QY 757 TACTCTGACATCGAATCAGAAGCATCTGGGACGAGAAGCTCTATCTGAAC 816
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3081 TACCGCTTCGCGTGGCATCCGTCAGTGCGAGTGAAGGCCAACAGTTCTGATTAAC 3140

QY 817 GGGAACTGTCTTTTGAAGGCTTTGAAAGCAGAGGAATCCCGTTCTGGGCGAG 876
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3141 CACAAACGTTCTACTTACTGCTTTGGTCTGTCATGAAGATGCGACTTACGTGGCAAA 3200

QY 877 GGCACCTTTTATCATTTGATGATAAAGCTTCAACCTTCTGAAGTGATCAACGGGAAT 936
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3201 GGATTCGATAACGTGCTGATGTGACAGCACGCAATTAATGACTGATTGGGCCAAC 3260

QY 937 TCTTTCAGGACCTTCACATCTCTTCACTGAGAGAGTGCTGATCTTCCGACAGACTC 996
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3261 TCCTACCGTACCTCGCATTTACCTTACGCTGAAGAGATGCTGACTGGGCAATGAACAT 3320

QY 997 GGAATCCTTGTGATAGAGGAAGCCCGGACGTTGGTATCAACAAGT----- 1042
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3321 GGCAATCGTGTGATGATGAACACTGCTGCTGCGCTTTTCTCTTTAGCATTTGGT 3380

QY 1043 -----ACCACTACAATCCCGAGACT 1062
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3381 TTGGAAGCGGCAACAAGCCGAAAGAACTGTACAGCGAAGAGGCAATCAACGGGAAACT 3440

QY 1063 CAGAAAGTAGCAGAAACAACATAAGAGAATGATGACAGACACAAGAACCATCCAGT 1122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3441 CAGCAAGCGCACTTACAGGCGATTAAAGAGCTGATAGCGCGTGAACAACCAACCAAGC 3500

QY 1123 GTGATCATGTGAGTGTGGCGAAGCAAGCAGAGTCCAAACATCCAGACGGAGGTTTC 1182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3501 GTGTGATGTGAGTATTGCCAACGAACCGGATACCGTCCGCAAGGTGCAACGGGATAT 3560

QY 1183 TTCAAGCCCTTTATGAGACTGCCAATGAATGATGGAACACGCCCCGTTGTCTATGTG 1242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 3561 TTGCGCCCACTGGCGGGAAGCAACGCGTAACTTCGACCCGACGCGTCCGATCACCTGCGTC 3620
QY 1243 AGCATGATGACGCAACGACGAGAAACAAGACGCTGCGCTGAAGTACTTCGACATC 1302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3621 AATGTAATGTTTCGCGACGCTTCACACCGATACCATCAGCGATCT-----CTTTGATGTG 3674

QY 1303 GTCTGTGTAACAGTACTACGCTGTGTACATCTATCAGGGAAGATAGAGAAGACTT 1362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3675 CTGTGCTGAAACCGTTATTACGATGTGTATGTCAAAGCGGCGATTGGAAACGCGACAG 3734

QY 1363 CAAGCTTGAAAAAGACATAGAAGCTCTATGCAAGGACAGAAAGCCCATCTTTGTC 1422
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3735 AAGGTAATGAAAAAGAACTTCTGGCCTGCGACGAGAAACTGCATCAGCCGATTATCATC 3794

QY 1423 ACAGAAATTCGGTGGGACGCGATAGCTGGCATCCATCAGATCCACTCAATGTCTCC 1482
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3795 ACCGAATACGGCGTGATACGTTAGCCGGCTGCACTCAATGTACACCGACATGTGAGT 3854

QY 1483 GAAGATACCAAGCAGAGCTGTTGAAAGACGATCAGGCTCCTTTGAAAAAAGACTAC 1542
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3855 GAAGATATCAGTGTGCATGCTGTGATATATCAACGCGCTTTGATCGCGTACGCGC 3914

QY 1543 ATCATCGAACAACAGTGTGGGCTTTGACAGATTTTAAAGCTCTCAGAAATGTGAAGA 1602
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3915 GTGCTGGTGAACAGGTATGAATTTCCGCAATTTTGGACCTGCAAGCAATTTGCGC 3974

QY 1603 CCCATTCTCAACCAAGGATGTTTCAACAAGACAGACAAACCAACTCGTTGCTCAT 1662
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3975 GTTGGCGGTAAACAAGAAAGGATCTTCACTCGGACCGCAACCGAAGTGGCGGCTTTT 4034

QY 1663 GTACTGAGAAGACTGTGAGTG 1684
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4035 CTGCTGCAAAAACGCTGACTG 4056

```

Search completed: September 1, 2005, 19:13:25
 Job time : 340 secs

This page Blank (usps)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 1, 2005, 17:43:49 ; Search time 1163 Seconds
(without alignments)
9510.845 Million cell updates/sec

Title: US-09-936-759-14

Perfect score: 1689

Sequence: 1 atggaagaccgcgaagaaa.....gaagactgtgagtgaggtc 1689

Scoring table:

IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 7338684 seqs, 327445616 residues

Total number of hits satisfying chosen parameters: 14677368

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

. Published Applications NA: *
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq: *
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq: *
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq: *
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq: *
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq: *
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq: *
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq: *
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq: *
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq: *
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq: *
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq: *
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq: *
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq: *
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq: *
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq: *
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq: *
19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq: *
20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq: *
21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq: *
22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq: *
23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq: *
24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq: *
25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq: *
26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1688	99.9	1689	17	US-10-364-649-14
2	1688	99.9	1692	17	US-10-369-493-26682
3	175.4	10.4	1926	22	US-10-757-093-1
4	174	10.3	2169	14	US-10-136-841-5
5	174	10.3	2169	17	US-10-272-531A-5
6	174	10.3	2169	17	US-10-272-483A-5
7	172.4	10.2	1956	17	US-10-421-175-1

8	172.4	10.2	2191	17	US-10-388-360-343	Sequence 343, App
9	172.4	10.2	2191	17	US-10-429-802-16	Sequence 16, Appl
10	172.4	10.2	2191	17	US-10-430-503-7	Sequence 7, Appl
11	172.4	10.2	2191	20	US-10-335-053-51	Sequence 51, Appl
12	166.8	9.9	2472	17	US-10-388-934-225	Sequence 225, App
13	159.6	9.4	1947	17	US-10-421-175-3	Sequence 3, Appl
14	141.6	8.4	1905	22	US-10-757-093-3	Sequence 7, Appl
15	141.4	8.4	1887	16	US-10-120-145-7	Sequence 45, Appl
16	141.4	8.4	3651	16	US-10-322-656-45	Sequence 48, Appl
17	141.4	8.4	4084	16	US-10-322-656-48	Sequence 5, Appl
18	137.6	8.1	1905	22	US-10-757-093-5	Sequence 27, Appl
19	135.4	8.0	1935	22	US-10-757-093-9	Sequence 9, Appl
20	130.4	7.7	1888	17	US-10-364-649-27	Sequence 108, App
21	126.2	7.5	1839	22	US-10-757-093-7	Sequence 20, Appl
22	114	6.7	3451	15	US-10-161-403-108	Sequence 109, App
23	114	6.7	3451	20	US-10-161-408-20	Sequence 109, App
24	114	6.7	3451	24	US-11-006-076-108	Sequence 109, App
25	114	6.7	14627	15	US-10-161-403-109	Sequence 109, App
26	114	6.7	14627	20	US-10-161-408-21	Sequence 21, Appl
27	114	6.7	14627	24	US-11-006-076-109	Sequence 13, Appl
28	112.4	6.7	1812	24	US-11-000-863-13	Sequence 24529, A
29	112	6.6	14683	13	US-10-109-853-1	Sequence 1, Appl
30	112	6.6	14683	19	US-10-817-950-1	Sequence 16, Appl
31	110.8	6.6	1809	19	US-10-432-777-16	Sequence 105, App
32	110.8	6.6	1812	15	US-10-161-403-105	Sequence 16, Appl
33	110.8	6.6	1812	17	US-10-369-493-24529	Sequence 33, Appl
34	110.8	6.6	1812	20	US-10-161-408-16	Sequence 105, App
35	110.8	6.6	1812	20	US-10-149-533A-33	Sequence 38, Appl
36	110.8	6.6	1812	24	US-11-006-076-105	Sequence 21, Appl
37	110.8	6.6	2001	18	US-10-239-907A-38	Sequence 29, Appl
38	110.8	6.6	2001	19	US-10-332-406A-21	Sequence 2, Appl
39	110.8	6.6	2001	19	US-10-800-161-29	Sequence 43, Appl
40	110.8	6.6	2186	20	US-10-755-328-2	Sequence 41, Appl
41	110.8	6.6	2725	18	US-10-239-907A-43	Sequence 5, Appl
42	110.8	6.6	2730	18	US-10-632-815-5	Sequence 49, Appl
43	110.8	6.6	3169	21	US-10-239-907A-49	Sequence 47, Appl
44	110.8	6.6	4072	18	US-10-239-907A-47	
45	110.8	6.6	4341	18	US-10-239-907A-47	

ALIGNMENTS

RESULT 1
US-10-364-649-14
; Sequence 14, Application US/10364649
; Publication No. US20030229921A1
; GENERAL INFORMATION:
; APPLICANT: Richard A. Jefferson and Jorge E. Mayer
; TITLE OF INVENTION: MICROBIAL B-GLUCURONIDASE GENES, GENE
; TITLE OF INVENTION: PRODUCTS, AND USES THEREOF
; FILE REFERENCE: 190106.405C1
; CURRENT APPLICATION NUMBER: US/10/364,649
; CURRENT FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: 10/364,649
; PRIOR FILING DATE: 2003-02-12
; PRIOR APPLICATION NUMBER: US 09/270,957
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1689
; TYPE: DNA
; ORGANISM: Thermotoga maritima
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(1689)
; OTHER INFORMATION: n = A,T,C or G
US-10-364-649-14

Query Match 99.9%; Score 1688; DB 17; Length 1689;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1689; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGTAAGACCGCAACGAAAGAGATTTATCTTATCTTGAATGGAGTTTGAAT 60
 Db 1 ATGTAAGACCGCAACGAAAGAGATTTATCTTATCTTGAATGGAGTTTGAAT 60
 OY 61 CTTGAAGTAACCGCAAGACAGACCAATCGCCGTTCTGGAAGCTGGAATGAGCAGTAC 120
 Db 61 CTTGAAGTAACCGCAAGACAGACCAATCGCCGTTCTGGAAGCTGGAATGAGCAGTAC 120
 OY 121 CAGGATCTGTGCTACGAAAGAGACCCCTTCACTACAAACCACTTCTAGCTTCCGAAG 180
 Db 121 CAGGATCTGTGCTACGAAAGAGACCCCTTCACTACAAACCACTTCTAGCTTCCGAAG 180
 OY 181 NAACTTTCACAAAAAACATCAGACTTTACTTGTGCGGTGAACACGGACTGCGAGGTC 240
 Db 181 NAACTTTCACAAAAAACATCAGACTTTACTTGTGCGGTGAACACGGACTGCGAGGTC 240
 OY 241 TTCTCAACGAGAGAAAGTGGAAGATCACAATGAATACCTTCCCTTGAAGTAGAT 300
 Db 241 TTCTCAACGAGAGAAAGTGGAAGATCACAATGAATACCTTCCCTTGAAGTAGAT 300
 OY 301 GTGACGGGAAAGTGAATCCGAGAGAAAGAACTCAAGGCTGTTGAGAAAGATTTG 360
 Db 301 GTGACGGGAAAGTGAATCCGAGAGAAAGAACTCAAGGCTGTTGAGAAAGATTTG 360
 OY 361 AAAGTGGGAGGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGAATTTT 420
 Db 361 AAAGTGGGAGGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACACCGTGGAATTTT 420
 OY 421 GGAAGTTTCCACCTGCAAACTTGAAGGTTCCCTTCCCTACGCTGAATCATAGGCTGTT 480
 Db 421 GGAAGTTTCCACCTGCAAACTTGAAGGTTCCCTTCCCTACGCTGAATCATAGGCTGTT 480
 OY 481 CTGATAGAGTTCAACAGCAAGCGAGAGATCTGACATCTGGTGAACACGAGTGAAGTCT 540
 Db 481 CTGATAGAGTTCAACAGCAAGCGAGAGATCTGACATCTGGTGAACACGAGTGAAGTCT 540
 OY 541 GAAACGGAGAAAGAACTTGGAAAAAGTGAAGTAGAAGTCTCAGAGAAGCGGTG 600
 Db 541 GAAACGGAGAAAGAACTTGGAAAAAGTGAAGTAGAAGTCTCAGAGAAGCGGTG 600
 OY 601 GGAACGAGAGATGACGATCAAACTTGGAGAGAGAGAAAAAGATTGAACATCCACAGA 660
 Db 601 GGAACGAGAGATGACGATCAAACTTGGAGAGAGAGAAAAAGATTGAACATCCACAGA 660
 OY 661 TTCTGCAAGGGAGTTTATCTCTGCAAAACGCCAGGTTCTGAGCCTCGAAGATCCATAT 720
 Db 661 TTCTGCAAGGGAGTTTATCTCTGCAAAACGCCAGGTTCTGAGCCTCGAAGATCCATAT 720
 OY 721 CTTTATCTCTCAAGGTGAACCTTGA AAAAGACGAGTACACTCTGACATCGGAATCAGA 780
 Db 721 CTTTATCTCTCAAGGTGAACCTTGA AAAAGACGAGTACACTCTGACATCGGAATCAGA 780
 OY 781 ACGATCAGCTGGGACGAGAGAGGCTCTATCTGAACGGGAAACCTGCTTTTGAAGGGC 840
 Db 781 ACGATCAGCTGGGACGAGAGAGGCTCTATCTGAACGGGAAACCTGCTTTTGAAGGGC 840
 OY 841 TTTGAAAAGCAGAGAAATTTCCCGCTTCTGGGGCAGGGCACCCTTTATCCATTGATGATA 900
 Db 841 TTTGAAAAGCAGAGAAATTTCCCGCTTCTGGGGCAGGGCACCCTTTATCCATTGATGATA 900
 OY 901 AAAAGCTTCAACCTTCTGAAGTGAATCAACGCGAATTTCTTCAGGACCTCTCATATCCT 960
 Db 901 AAAAGCTTCAACCTTCTGAAGTGAATCAACGCGAATTTCTTCAGGACCTCTCATATCCT 960
 OY 961 TACAGTGAAGAGTGGCTGGATCTTGCCGACAGACTCGGAATCTTGTGATAGACGAAGCC 1020
 Db 961 TACAGTGAAGAGTGGCTGGATCTTGCCGACAGACTCGGAATCTTGTGATAGACGAAGCC 1020
 OY 1021 CCGCAGCTTGGTATACAAAGGTACCACTAACATCCCGAGACTCAGAAAGATAGCAGAAAGAC 1080
 Db 1021 CCGCAGCTTGGTATACAAAGGTACCACTAACATCCCGAGACTCAGAAAGATAGCAGAAAGAC 1080

OY 1081 AACATAGAAGATGATCGACAGACACAAAGAACCATCCAGTGTGATCATGTGAGTGTG 1140
 Db 1081 AACATAGAAGATGATCGACAGACACAAAGAACCATCCAGTGTGATCATGTGAGTGTG 1140
 OY 1141 GCGAAGCAACCAAGTCCCAACCATCCAGACGGCGGAGGTTCTTCAAAAGCCCTTTATGAG 1200
 Db 1141 GCGAAGCAACCAAGTCCCAACCATCCAGACGGCGGAGGTTCTTCAAAAGCCCTTTATGAG 1200
 OY 1201 ACTGCCAATGAATGGATCGAACACAGCCCGCTTGTCTATGTGAGCATGATGACGCCACCA 1260
 Db 1201 ACTGCCAATGAATGGATCGAACACAGCCCGCTTGTCTATGTGAGCATGATGACGCCACCA 1260
 OY 1261 GACGAGAGAACAGAGACGTTGGCGCTGAAGTACTTCGACATCTGTGTGAACAGGTAC 1320
 Db 1261 GACGAGAGAACAGAGACGTTGGCGCTGAAGTACTTCGACATCTGTGTGAACAGGTAC 1320
 OY 1321 TACGGCTGTATCATCTATCAGGGAAGATGAGAAAGACTTCAAGCTCTGGA AAAAGAC 1380
 Db 1321 TACGGCTGTATCATCTATCAGGGAAGATGAGAAAGACTTCAAGCTCTGGA AAAAGAC 1380
 OY 1381 ATAGAAGACTCTATGCAAGGCAAGAAAGCCCATCTTGTGCAAGAAATTCGTTGCGGAC 1440
 Db 1381 ATAGAAGACTCTATGCAAGGCAAGAAAGCCCATCTTGTGCAAGAAATTCGTTGCGGAC 1440
 OY 1441 GCGATAGCTGGCATCCACTAGATCCACCTGAATGTTCTCGAAGAGTACCAAGCAGAG 1500
 Db 1441 GCGATAGCTGGCATCCACTAGATCCACCTGAATGTTCTCGAAGAGTACCAAGCAGAG 1500
 OY 1501 CTCGTTGAAAAAGACGATCAGGCTCTTTGAAAAAAGACTACATCATCGGAACACACGTG 1560
 Db 1501 CTCGTTGAAAAAGACGATCAGGCTCTTTGAAAAAAGACTACATCATCGGAACACACGTG 1560
 OY 1561 TGGGCTTTGAGATTTTAAAGCTCCTCAGATGTGAGAAAGCCCATCTCAACCAAG 1620
 Db 1561 TGGGCTTTGAGATTTTAAAGCTCCTCAGATGTGAGAAAGCCCATCTCAACCAAG 1620
 OY 1621 GGTGTTTCAAGAGACAGACCAACCCAACTCGTTGCTCATGTACTGAGAAAGCTGTGG 1680
 Db 1621 GGTGTTTCAAGAGACAGACCAACCCAACTCGTTGCTCATGTACTGAGAAAGCTGTGG 1680
 OY 1681 AGTGAGTT 1689
 Db 1681 AGTGAGTT 1689

RESULT 2
 US-10-369-493-26682
 ; Sequence 26682, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 26682
 ; LENGTH: 1692
 ; TYPE: DNA
 ; ORGANISM: Thermotoga maritima
 US-10-369-493-26682
 Query Match 99.9%; Score 1688; DB 17; Length 1692;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1688; Conservative 0; Mismatches 1; Indels 0; Gaps 0;


```
QY 1 ATGTTAAGACCCGAAACGAAACAGAAGATTTATTCTTATCTTGAATGAGTTTGAAAT 60
Db 1 ATGTTAAGACCCGAAACGAAACAGAAGATTTATTCTTATCTTGAATGAGTTTGAAAT 60
QY 61 CTGGAAGTAAACCGAAAGACAGACCAATCGCCGTTCTGGAAGCTGGAATGAGCAGTAC 120
Db 61 CTGGAAGTAAACCGAAAGACAGACCAATCGCCGTTCTGGAAGCTGGAATGAGCAGTAC 120
QY 121 CAGGATCTGTGCTACGAAAGAGACCCTTCACTTCAAAAACACCTTCTACGTTCCGAAG 180
Db 121 CAGGATCTGTGCTACGAAAGAGACCCTTCACTTCAAAAACACCTTCTACGTTCCGAAG 180
QY 181 NAACTTCACAAAAACACATCAGACTTTACTTGTGCGGTGAACACGGACTGCGAGTTC 240
Db 181 GAACTTCACAAAAACACATCAGACTTTACTTGTGCGGTGAACACGGACTGCGAGTTC 240
QY 241 TTCTTCAACGAGAGAAAGTGGAGAGATCACAATTGAATACCTTCCCTTGAAGTAGAT 300
Db 241 TTCTTCAACGAGAGAAAGTGGAGAGATCACAATTGAATACCTTCCCTTGAAGTAGAT 300
QY 301 GTGACGGGGAAAGTAAATCCGAGAGAACGAACTCAGGGTGGTGTGAGAACAGATTG 360
Db 301 GTGACGGGGAAAGTAAATCCGAGAGAACGAACTCAGGGTGGTGTGAGAACAGATTG 360
QY 361 AAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACACGGTGGATTTT 420
Db 361 AAAGTGGAGAGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACACGGTGGATTTT 420
QY 421 GGAAGTTTCCACCTGCAAACTTGCATCTTCCCTACGGTGAATCATTAAGGCTGTT 480
Db 421 GGAAGTTTCCACCTGCAAACTTGCATCTTCCCTACGGTGAATCATTAAGGCTGTT 480
QY 481 CTGATAGAGTTCACAGACCAACGCGAGGATCTGCACATCTGGGTGACACGAGTGAATCT 540
Db 481 CTGATAGAGTTCACAGACCAACGCGAGGATCTGCACATCTGGGTGACACGAGTGAATCT 540
QY 541 GAACCGGAGAAAGAACTTGAAAAAGTGAAGTAGAAGTCTCGAAAGACGGTG 600
Db 541 GAACCGGAGAAAGAACTTGAAAAAGTGAAGTAGAAGTCTCGAAAGACGGTG 600
QY 601 GGACAGAGATGACGATCAAACTTGAGAGAGAAAGAAAGATTAGACATCCAACAGA 660
Db 601 GGACAGAGATGACGATCAAACTTGAGAGAGAAAGAAAGATTAGACATCCAACAGA 660
QY 661 TTCGTGAAAGGGAGTTGATCTCTGAAAACGCCAGTTCTGGAGCCTCGAAGATCCATAT 720
Db 661 TTCGTGAAAGGGAGTTGATCTCTGAAAACGCCAGTTCTGGAGCCTCGAAGATCCATAT 720
QY 721 CTTTATCTCTCAAGTGAACCTTGAAAAAGACGAGTACACTCTGACATCGGAATCAGA 780
Db 721 CTTTATCTCTCAAGTGAACCTTGAAAAAGACGAGTACACTCTGACATCGGAATCAGA 780
QY 781 ACGATCAGCTGGGACGAGAAAGGCTTAATCTGAACGGGAAAACCTGTCTTTTGAAGGCG 840
Db 781 ACGATCAGCTGGGACGAGAAAGGCTTAATCTGAACGGGAAAACCTGTCTTTTGAAGGCG 840
QY 841 TTTGAAAACGACGAGGAATTCCTCCGTTCTGGGGCAGGGCACTTTTATCCATTGATATA 900
Db 841 TTTGAAAACGACGAGGAATTCCTCCGTTCTGGGGCAGGGCACTTTTATCCATTGATATA 900
QY 901 AAAGACTTCAACCTTCTGAAGTGAATCAACGGAATCTTTCAGGACCTCTCACTATCT 960
Db 901 AAAGACTTCAACCTTCTGAAGTGAATCAACGGAATCTTTCAGGACCTCTCACTATCT 960
QY 961 TACAGTGAAGAGTGGCTGGATCTTGCCGACAGACTCGGAATCCTTGATAGACGAAGCC 1020
Db 961 TACAGTGAAGAGTGGCTGGATCTTGCCGACAGACTCGGAATCCTTGATAGACGAAGCC 1020
QY 1021 CCGCAGCTTGGTATCACAGGTACCACTAACAATCCCGAGACTCAGAAAGATGACAGAAAC 1080
Db 1021 CCGCAGCTTGGTATCACAGGTACCACTAACAATCCCGAGACTCAGAAAGATGACAGAAAC 1080
QY 1081 AACATAAGAAAGATGATGACAGACACAGAACCATCCAGTGTGATCATGTGAGTGTG 1140
```

```
Db 1081 AACATAAGAAATGATCGACAGACACAAGAACCATCCCAAGTGTGATCATGTGAGTGTG 1140
QY 1141 GCGAAGAACCCAGAGTCCCAACCATCCAGACGCGGAGGTTTCTTCAAAGCCCTTTATGAG 1200
Db 1141 GCGAAGAACCCAGAGTCCCAACCATCCAGACGCGGAGGTTTCTTCAAAGCCCTTTATGAG 1200
QY 1201 ACTGCCAATGAATGATCGAACACGCCCCGTTGTATGTGTGAGCATGATGACGCCACCA 1260
Db 1201 ACTGCCAATGAATGATCGAACACGCCCCGTTGTATGTGTGAGCATGATGACGCCACCA 1260
QY 1261 GACGAGAGAACCAAGAGACGTGGCGCTGAAGTACTTTCGACATCGTCTGTGTAACAGTAC 1320
Db 1261 GACGAGAGAACCAAGAGACGTGGCGCTGAAGTACTTTCGACATCGTCTGTGTAACAGTAC 1320
QY 1321 TACGGCTGTACATCTATCAGGGAAGATAGAAGAGACTTCAAGCTCTGAAAAAGAC 1380
Db 1321 TACGGCTGTACATCTATCAGGGAAGATAGAAGAGACTTCAAGCTCTGAAAAAGAC 1380
QY 1381 ATAGAAGAGCTCTATGCAAGGCACAGAAAGCCCATCTTTGTACAGAAATTCGGTCCGAC 1440
Db 1381 ATAGAAGAGCTCTATGCAAGGCACAGAAAGCCCATCTTTGTACAGAAATTCGGTCCGAC 1440
QY 1441 GCGATAGCTGGCATCTCACTACGATCACTCAATGTCTCCGAAAGATACCAAGCAGAG 1500
Db 1441 GCGATAGCTGGCATCTCACTACGATCACTCAATGTCTCCGAAAGATACCAAGCAGAG 1500
QY 1501 CTCGTTGAAAAAGACGATCAGGCTCCTTTGAAAAAAGACTACATCATCGGAACACAGTG 1560
Db 1501 CTCGTTGAAAAAGACGATCAGGCTCCTTTGAAAAAAGACTACATCATCGGAACACAGTG 1560
QY 1561 TGGGCTTTGACAGATTTTAAGACTCTCTCAGAAATGAGAAAGACCATTTCTCAACCAAG 1620
Db 1561 TGGGCTTTGACAGATTTTAAGACTCTCTCAGAAATGAGAAAGACCATTTCTCAACCAAG 1620
QY 1621 GGTGTTTTCACAAAGAGACAGACAAACCCAACTCGTGTCTCATGTACTGAGAAAGCTGTG 1680
Db 1621 GGTGTTTTCACAAAGAGAGACAGACAAACCCAACTCGTGTCTCATGTACTGAGAAAGCTGTG 1680
QY 1681 AGTGAGGTT 1689
Db 1681 AGTGAGGTT 1689

RESULT 3
US-10-757-093-1
; Sequence 1, Application US/10757093
; Publication No. US20050153448A1
; GENERAL INFORMATION:
; APPLICANT: CAMBIA
; TITLE OF INVENTION: Fungal beta-glucuronidase genes and gene products
; FILE REFERENCE: 415
; CURRENT APPLICATION NUMBER: US/10/757,093
; CURRENT FILING DATE: 2004-01-14
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Scopulariopsis sp. isolate RP38.3
US-10-757-093-1

Query Match 10.4%; Score 175.4; DB 22; Length 1926;
Best Local Similarity 49.0%; Pred. No. 3.2e-45;
Matches 777; Conservative 0; Mismatches 717; Indels 92; Gaps 8;
```

```
QY 292 GAAATGATGTGACGGGGAAAGTAAATCCGGAGAGAACGAATCTCAGGGTGTGTGAG 351
    |||||
Db 442 GAGGTGACCTGACCGACCTTGTGTGCGCCGGAGAGAGTTCCGCTGACTGTGTGTGTC 501
QY 352 AAC-AGATTGAAAGTGGAGGATTTCCCTCGAAGGTTCCAGACAGCGGCACTCACACCGT 410
    |||||
Db 502 AACAAATATCTGACTTGGCAGACCAATCCCGCTGTGAGGTGTGACCAACGAGCGTGTG 561
QY 411 GGGATTTTGGAAAGTTTCCACCCTGCAAACTTCGACTTCTTCCCTACCGGTGATCAT 470
    |||||
Db 562 AAGCTTCGACAGGACTAC-----AACCAAGACTTCTAACAACTACGCTGGAATTGC 611
QY 471 AAGCCTGTCTGATAGAGTTCAACAGACCAAGGAGTACTCGACATCTGGTGACAC 530
    |||||
Db 612 ACGTTCGCTCTGCTAATCTCCGTGCTGATGTTCATGTGAGCAAGTCACTGT--AC 668
QY 531 GAGTGTCTGAAACCGAGAGAAACTTGGAAAAGTGAAGTGAAGTGAAGTCTCAGA 590
    |||||
Db 669 TACCGAAGAACGACGAGAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCGG 728
QY 591 AGAAGCGGTGGACAGAGATGACGATCAAACTTGGAGAGAGAGAGAGAGAGAGATTAAG 650
    |||||
Db 729 GTCTAACGACACTCAGGCTAGGGTCACTTGTATGATGAGAGAGAGAGAGAGAGAGAG 788
QY 651 ATCCAACAGATTCTGCGAAGGAGGAGTCACTCTCGAAGAGCCAGGTTCTGAGCCTGA 710
    |||||
Db 789 GGCATCGGAGCTG---GAGGGAGCTTGAACGTGAGCCCGTGAATCTCTGGCAGCCGGG 845
QY 711 AGATCATATCTTTATCTCTCAAGGTGGAACCTTGAAGAGAGAGAGAGAGAGAGAGAG 769
    |||||
Db 846 CGCGGCGTACCTCTTACACTCTTCCGCTTGAACCTCTTTCGAGAGATACGCTGTCACAC 905
QY 770 -----TCGGAATCAGAACGATCAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAG 815
    |||||
Db 906 TTATGATTTACCGGTTGGTGTACGGTCCGTTAGGTTGAAGAGAACAGTTCTCTATCAA 965
QY 816 CGGAAACCTGTCTTTTGAAGGGCTTGGAAAGCAGAGAAATCCCGTTCTGGGGCA 875
    |||||
Db 966 CGGCAAGCCCTTCTACTTCAACCGGCTTGGCAAGCAGAGACAGAGAGAGAGAGAGAG 1025
QY 876 GGGACCTTTTATCATGATGATTAAGACTTCAACTTCTGAAGTGAATCAAGCGGAA 935
    |||||
Db 1026 GGGCTACGACCGGCTACATGATCATGATTTGAGCTCATGAGTGGAGCGGCGCAA 1085
QY 936 TTCTTCAGGACCTCTCACTATCTTCAAGTGAAGAGTGGCTGATCTTCCGACAGACT 995
    |||||
Db 1086 CTCTTCGAGACCTCCCACTACCCCTACGCGGAGGTCATGAGTACGCCAGCGTCA 1145
QY 996 CGGAATCCTTGATAGACGAAGCCCGCAGCTGGTAT----- 1034
    |||||
Db 1146 CGGCATCGTGTGATGACGAGGTGCGCGCGCTGCTGAACCTGGGCATCAGCGCAGG 1205
QY 1035 -----CACAGGTACCACTACATCCCGAGACTCA 1064
    |||||
Db 1206 CCTCAGGGGAGATGAGCCCGCCAGACCTTACGAGAGAGAGTTAAACAACGAGACGCA 1265
QY 1065 GAAATAGCAGAGACAAATAGAGATGATCGACAGACAGAGAACCATCCAGTGT 1124
    |||||
Db 1266 AAAAGACACGCGCCGCTCCGTGATGATCCACCGTGAACAAGAACCAAGCGCTCCGT 1325
QY 1125 GATCATGTGAGTGTGCGAAGCAAGACAGAGTCCCAACCATCCAGACGCGGAGGTTCTT 1184
    |||||
Db 1326 TGTCACTGTGTGCTACCAACGAGCGCGCTCCGCGAGGACGCGGAGTACTT 1385
QY 1185 CAAAGCCCTTTATGAGACTGCCAATGAATGATGATGAGAACAGCGCCCGTTGTCAATGTGAG 1244
    |||||
Db 1386 CCAAGCCCTGTGCTGAGCTAACCCGCGAGCTGGAACCCCAACCGCGCTCACTTCAACCA 1445
QY 1245 CATGATGACGACCAAGACGAGAGAACAGAGAGAGAGAGAGAGTGAAGTACTTCAGATCGT 1304
    |||||
Db 1446 CGTATGGGCGCACCGGTGACAAAGTGCCTCATCTCGATCT-----TTTCGACTTCT 1499
```

```
QY 1305 CTGTGTGAACAGTACTACGGCTGTGATCTATCAAGGAAAGATAGAAAGAGACTTCA 1364
    |||||
Db 1500 TTCTCTCAACCGCTACTACGGGTGTGATCGTCCAAACGGGCGACCTGAGTCCGCGAGGT 1559
QY 1365 AGCTCTGAAAGAACATAGAGAGCTCTATGCAAGGACACAGAAAGCCCATCTTTGTCA 1424
    |||||
Db 1560 CGCCATGAGAGAGAGAGCTCTCTCCAGTGGGTGCAAGATATGACAAAGCCTATCATGTTC 1619
QY 1425 AGAATTCGGTGGGACCGCGATAGCTGGCATCCACTACGATCCACCTCAATGTCTCCGA 1484
    |||||
Db 1620 CGAGTACGGCGCCGACACCCCTGGCCGGTCTTCCACGCGGTGCAAGAGTGTGTGTCGA 1679
QY 1485 AGAGTACCAAGCAGAGCTCGTTGAAAAAGCATCAAGGCTCCTTTGAAAAAGACTACAT 1544
    |||||
Db 1680 GGAATACCAAGCAACCTCTCGCGCATGTGCAACAAGGTCTTTGACAGCATGACTCCAT 1739
QY 1545 CATCGAACAACAGCTGTGGGCTTTGACAGATTTTAAGACTCTCTCAGA--ATGTGAGAG 1601
    |||||
Db 1740 TGTGGCGAGCAGTGTGGAATTTGCTGATTTCCAGACTCTCATACTGTGTCAACCG 1799
QY 1602 ACCCATTTCAACCAAGGAGGTTTTCACAGAGACAGACAGAACCCAACTGTTGCTCA 1661
    |||||
Db 1800 TGTGTATGAAACAAAGAGGTTGTTTACCGGTGAGCGGAGGCTTAAGGCGCGGACACA 1859
QY 1662 TGTACTGAGAAAGCTGTGAGTGAAG 1687
    |||||
Db 1860 TGAGCTCAAGAGGCGGTGCTTGACG 1885
```

RESULT 4

```
US-10-136-841-5
; Sequence 5, Application US/10136841
; Publication No. US20030082176A1
; GENERAL INFORMATION:
; APPLICANT: Lebowitz, Jonathan
; APPLICANT: Beverley, Stephen
; TITLE OF INVENTION: SUBCELLULAR TARGETING OF THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-007
; CURRENT APPLICATION NUMBER: US/10/136,841
; CURRENT FILING DATE: 2002-08-22
; PRIOR APPLICATION NUMBER: US 60/287,531
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: US 60/304,609
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/329,461
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/351,276
; PRIOR FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 2169
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A recombinant sequence incorporating a signal peptide sequence,
; OTHER INFORMATION: the mature human beta-glucuronidase sequence, a bridge of three
; OTHER INFORMATION: amino acids, and an IGF-II sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2166)
; US-10-136-841-5
```

Query Match 10.3%; Score 174; DB 14; Length 2169;
Best Local Similarity 52.2%; Pred. No. 9.8e-45;
Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;

```
QY 737 TGGAACTTGAAAAAGACGAGTACACTCTGACATCGGAATCAGAAAGATCAGCTGGAGC 796
    |||||
Db 941 TGGGGCTGTGTCTGACTTCTACACACTCCGTGTGGGATCCGCACTGTGCTGTCAACCA 1000
    |||||
QY 797 AGAAGAGGCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGGCTTTGAAAGCACAGAG 856
    |||||
```

```
Db 1001 AGAGCCAGTTCCTCATCATGGGAACCTTTCTATTTCCACGGGTGTCACAAGCATGAGG 1060
QY 857 AATCCCCCTTCTGGGCGAGGGACCTTTTATCCATTGATGATAAAGACTTCAACTTC 916
Db 1061 ATGCGGACATCCGAGGGAAGGGCTTGCATGCGCCGCTGCTGTGAAGCACTTCAACTGC 1120
QY 917 TGAAGTGGATCAACGGGAATTCTTTGAGGACCTCTCACTATCCTTACAGTGAAGAGTGGC 976
Db 1121 TTCGCTGGCTTGGTGCCAACGCTTTCCTGTAACGACCTAACCTTATGACAGAGAACTGA 1180
QY 977 TGGATCTTGGCCGACAGACTCGGAATCTTGTGATAGCGAAGCCCGGACGTTGGTATCA 1036
Db 1181 TGCAGATGTGTGACCGCTATGGGATTGTGTCATGATGAGTGTCCCGGCGTGGGTG 1240
QY 1037 CAAGGTACCA--CTACAATCCCGAGACTCAGAAAGATAGCAGAAACAACATAAGAGAA 1093
Db 1241 CGTGGCCGAGTTCTTCAACAACGTTTCTGTGATCAACATGAGGTGATGAAAG 1300
QY 1094 TGATCGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGCAACAG 1153
Db 1301 TGTGCGTAGGAGCAAGAACCAACCGCGGTGATGATGTGTGTGGCCAAACGAGCTTG 1360
QY 1154 AGTCCAAACCATCCAGACGCGGAGGTTTCTTCAAAAGCCCTTATGAGACTGCCAATGAAA 1213
Db 1361 CGTCCACCTAGAACTGTCTGTGCTACTTGAAGATGTGATCGTCAACCAAAATCCT 1420
QY 1214 TGGATCGAAACACGCCCCGTGTGCTATGATGAGCATGATGAGCGCACAGAGAGAA 1273
Db 1421 TGGACCCCTCCCGGCTGTGACCTTGTGAGCAACTTAATAATGAG----- 1468
QY 1274 GAGACGTGGCGCTGAAGTACTTTCGATCGTGTGTGAACAGGTACTACGCTGGTACA 1333
Db 1469 CAGACAAGGGGCTCGCTATGTGATGTGATCTGTTGAACAGCTACTCTTGGTATC 1528
QY 1334 TCTATCAGGGAAGATAGAAGAGACTTCAAGCTGTGAAAAAGACATAGAAGACTCT 1393
Db 1529 ACGACTACGGGCACTGGAAGTGTGATGACTGACGTGGCCACCGATTGAGAACTGGT 1588
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTGTGTCAAGAAATTCGGTGGCGAGCGATAGTGCA 1453
Db 1589 ATAGAAGTATCAGAAGCCCATTAATCAGAGCGAGTATGAGCAGAAACGATTGACGGGT 1648
QY 1454 TCCACTACGATCCACTCAAAATGTCTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAAG- 1512
Db 1649 TTCAACCAAGATCCACTCTGATGTCTGATGAGTACCAAGAAAGTCTGCTAGAGCAGT 1708
QY 1513 ACGATCAGGCTCCTTTGAAAAA-----AGACTACATCATCGGAACACACGTGGGCT 1567
Db 1709 ACCATCTGGGTCTGATCAAAAAACGAGAAATATGTGTTGAGAGCTCATTTGGAATT 1768
QY 1568 TTGCAGATTTTAAGACTCCTCAGAAATGTGAGAGAGACCATTTCTCAACCAACAAGGTTT 1627
Db 1769 TTGCGGATTTTCATGATGAACAGTCAACCGACGAGAGTGTGGGAAATAAAGGGGATCT 1828
QY 1628 TCACAAGAGACAGAACCAACCAACTGCTGTCTCATGTACTGAGA 1671
Db 1829 TCACTCGGACAGAGAACCAAAAGTCAAGCGTCTTTTGGCA 1872
```

RESULT 5

```
US-10-272-531A-5
; Sequence 5, Application US/10272531A
; Publication No. US20040005309A1
; GENERAL INFORMATION:
; APPLICANT: Lebowitz, Jonathan H
; APPLICANT: Beverly, Stephen
; APPLICANT: Sly, William S.
; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-009
; CURRENT APPLICATION NUMBER: US/10/272,531A
; CURRENT FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/384,452
; PRIOR FILING DATE: 2002-05-29
```

```
; PRIOR APPLICATION NUMBER: US 60/386,019
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/408,816
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2169
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A recombinant DNA sequence incorporating a signal peptide sequenc
; OTHER INFORMATION: e, the mature human beta-glucuronidase sequence, a bridge of thre
; OTHER INFORMATION: e amino acids, and an IGF-II sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2166)
; OTHER INFORMATION:
US-10-272-531A-5
```

Query Match 10.3%; Score 174; DB 17; Length 2169;
Best Local Similarity 52.2%; Pred. No. 9.8e-45;
Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;

```
QY 737 TGGAACTTGAAAAAGACGATACACTCTGACATCGGAATCAGAACGATCAGCTGGGACG 796
Db 941 TGGGCGCTGTGTGCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGCTGTACCA 1000
QY 797 AGAAGAGCTCTATCTGAACGGGAACCTGTCTTTTGAAGGCTTTGAAAGCAGAG 856
Db 1001 AGAGCCAGTTCCTCATCATGGAACCTTCTATTTCCACGGGTGTCACAAGCATGAGG 1060
QY 857 AATCCCCCTTCTGGGCGAGGGACCTTTATCCATTGATGATAAAGACTTCAACTTC 916
Db 1061 ATGCGGACATCCGAGGGAAGGGCTTGCATGCGCCGCTGCTGTGAAGACTTCAACTGC 1120
QY 917 TGAAGTGGATCAACGGGAATTCTTCAAGACCTCTCACTATCCTTACAGTGAAGAGTGGC 976
Db 1121 TTCGCTGGCTTGGTGCCAACGCTTTCCTGATCAACGACCTATGACAGAGAACTGA 1180
QY 977 TGGATCTTGGCAGAGACTCGGAATTCCTTGTGATAGAAGAGCCCGACGTTGGTATCA 1036
Db 1181 TGCAGATGTGTGACCGCTATGGAATGTGTGATGATGATGATGTGTCAGTGTGCGGCTGG 1240
QY 1037 CAAGGTACCA--CTACAATCCGAGACTCAGAAAGATAGCAGAAACAACATAGAAGAA 1093
Db 1241 CGTCCCGGAGTCTTCAACAACGCTTCTGTGATCAACCATGAGTATGAGAAG 1300
QY 1094 TGATCGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGAAGCAACAG 1153
Db 1301 TGTGCGTAGGAGCAAGAACCAACCGCGGTGATGATGTGATGTGATGTGCGCAACGAGCTG 1360
QY 1154 AGTCCAAACCATCCAGACGCGGAGGTTTCTTCAAAAGCCCTTATGAGACTGCCAATGAAA 1213
Db 1361 CGTCCACCTAGAACTGTCTGCTACTACTTGAAGATGTGATGCTCACACCAATCCT 1420
QY 1214 TGGATCGAAACAGCGCCCTGTGTGATGAGCATGATGAGCGCACAGAGAGAAACA 1273
Db 1421 TGGACCCCTCCCGGCTGTGACCTTTTGTGAGCAACTTAACCTATGACG----- 1468
QY 1274 GAGACGTGGCGCTGAAGTACTTTCGACATCGTGTGTGAACAGGTACTACGGCTGTACA 1333
Db 1469 CAGACAAGGGGCTCCGCTATGTGATGTGATGTGATGTGATGTGATGTGATGTGATG 1528
QY 1334 TCTATCAGGGAAGATAGAAGAGACTTCAAGCTCTGAAAAAGACATAGAAGACTCT 1393
Db 1529 ACGACTACGGGCACTGGAAGTGTGATGAGTGTGAGCTGTGGCCACCGATTGAGAACTGGT 1588
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTGTGTCAACAGAAATTCGGTGGCGAAGCGATAGTGCCA 1453
Db 1589 ATAGAAGTATCAGAAGCCCATTAATTCAGAGCGAGTATGAGAGCAAGAAAGATTGACAGGGT 1648
QY 1454 TCCACTACGATCCACTCAAAATGTCTCCGAAGAGTACCAAGCAGAGCTCGTTGAAAAG- 1512
```

```
Db      1649 TTCAACGAGATCCACTCTGATGTTCACTGAAGAGTACAGAAAAGTCTGTAGAGCAGT 1708
QY      1513 ACGATCAGGCTCCTTTTGA AAAA-----AGACTACATCATCGGAACACAGTGTGGCCT 1567
Db      1709 ACCATCTGGGTCTGANTCAAAAACGAGAAAATATGTGTTGAGAGCTCATTTGGAATT 1768
QY      1568 TTGCAGATTTTAAGACTCCTCAGAAATGTAGAAAGCCATTCTCAACCAAGGCTGTTT 1627
Db      1769 TTGCCGATTTATGATGATGAACAGTCAACCGACGAGAGTGTCTGGGAATAAAAAGGGATCT 1828
QY      1628 TCACAAGAGACAGACAAACCAACTGCTGTCTCATGTACTGAGA 1671
Db      1829 TCACTCGGACAGACAAACCAAAAGTGACGCTTCCTTTGCGA 1872
```

RESULT 6

```
US-10-272-483A-5
; Sequence 5, Application US/10272483A
; Publication No. US2004006008A1
; GENERAL INFORMATION:
; APPLICANT: Lebowitz, Jonathan H
; APPLICANT: Beverley, Stephen
; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-007CP
; CURRENT APPLICATION NUMBER: US/10/272,483A
; CURRENT FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/287,531
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: US 10/136,841
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/384,452
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 60/386,019
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/408,816
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 60/304,609
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/329,461
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/351,276
; PRIOR FILING DATE: 2002-01-23
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 2169
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A recombinant DNA sequence incorporating a signal peptide sequenc
; OTHER INFORMATION: e, the mature human beta-glucuronidase sequence, a bridge of thre
; OTHER INFORMATION: e amino acids, and an IGF-II sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2166)
; OTHER INFORMATION:
US-10-272-483A-5
```

Query Match 10.3%; Score 174; DB 17; Length 2169;
Best Local Similarity 52.2%; Pred. No. 9.8e-45;
Matches 493; Conservative 0; Mismatches 430; Indels 21; Gaps 4;

```
QY      737 TGAACCTGAAAAAGAGAGTACACTGTGACATCGGAATCAGACATCAGCTGGAGC 796
Db      941 TGGGGCTGTGTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGCTGTACCA 1000
QY      797 AGAAGAGGCTCTATCTGAACGGGAACCTGTCTTTTGAAGGCTTTGGAAGCAGAG 856
Db      1001 AGAGCCAGTTCCTCATCATGGAACCTTTCTATTCCACGGTGTCAACAAGCATGAG 1060
QY      857 AATTCCCGCTTCTGGGGCAGGGCACTTTTATTCATGTGATATAAAGACTTCAACTTC 916
```

```
Db      1061 ATCCGACATCCGAGGGAAGGCTTCGACTGCGCCGCTGCTGGTGAAGACTTCAACCTGC 1120
QY      917 TGAAGTGATCAACGCCAATCTTTACAGAGCTCTCACTATCTTACAGTGAAGAGTGCC 976
Db      1121 TTCCTGGCTGTGTGCCAACGCTTTCCTGATCCAGCCACTACCCCTATGACAGGAAGTGA 1180
QY      977 TGGATCTTGCCCAACAGACTCGGAATCCTTGTGATAGACGAAGCCCGCAGCTGGTATCA 1036
Db      1181 TGCAAGATGTGTGACCGCTATATGGGATTTGTGTGATCATCGATGAGTGTCCCGCGTGGTCTGG 1240
QY      1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAAAGATAGCAGAAAGACAACATTAAGAAGA 1093
Db      1241 CGCTGCCGAGTTCTTCAACAAGTTCTCTGCAATCACACATGCAAGGTGATGGAAGAAG 1300
QY      1094 TGATCGACAGACACAGAACCATCCCAAGTGTGATCATGTGAGTGTGGCGAACCAG 1153
Db      1301 TGGTGCATAGGACACAGAACCACCCCGCGTGTGATGTGTGTGGCCAACGACCTG 1360
QY      1154 AGTCCAACCATCCAGACGCGGAGGGTTCTTCAAAAGCCCTTATGAGACTGCCAATGAAA 1213
Db      1361 GGTCCACCTTAGAATCTGCTGGCTACTACTTGAAGATGTGTGCTCACACCAATCCT 1420
QY      1214 TGGATCGAACAGCCCCGTTGTTCATGTGTGATGATGATGAGCGCAGACGAGAACAA 1273
Db      1421 TGAACCCCTCCCGGCTGTGACTTGTGAGCAACTTAACATATGCAG----- 1468
QY      1274 GAGACGTGGCGCTGAAGTACTTTCGACATCGTCTGTGTGAACAGTACTACGGCTGTACA 1333
Db      1469 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTGAACAGCTTACTACTCTTGGTATC 1528
QY      1334 TCTATCAGGGAAGATGAGAAGAGACTTCAAGCTCTGGA AAAAGACATAGAAAGCTCT 1393
Db      1529 ACGACTACGGGCACTTGAGTTGATTCACTGCAAGCTGGCCACCCAGTTTGAGAACTGGT 1588
QY      1394 ATGCAAGGCACAGAAAGCCCATCTTTGTCAAGAAATTCGGTGGGACGCGATAGCTGGCA 1453
Db      1589 ATAAGAAGTATCAGAAAGCCCATTTATTACAGCGAGTATGAGCAGAAACGATTGCAAGGT 1648
QY      1454 TCCACTACGATCCCACTCAAAATGTTCTCCGAAGAGTACCAGCAGAGCTCGTTGAAAAAG- 1512
Db      1649 TTCACCAAGATCCCACTCTGATGTTCACTGAAGAGTACCAGAAAAGTCTGCTAGAGCAGT 1708
QY      1513 ACGATCAGGCTCCTTTTGA AAAA-----AGACTACATCATCGGAACACACGTTGGGCCT 1567
Db      1709 ACCATCTGGGTCTGATCAAAAACGAGAAAATATGTGTTGAGAGCTCATTTGGAATT 1768
QY      1568 TTGCAGATTTTAAGACTCCTCAGAAATGTGAGAAGACCCATTCTCAACCAAGGCTGTTT 1627
Db      1769 TTGCCGATTTATGATGATGAACAGTCAACCGACGAGAGTGTCTGGGAATAAAAAGGGATCT 1828
QY      1628 TCACAAGAGACAGACAAACCAACTGCTGTGCTCATGTACTGAGA 1671
Db      1829 TCACTCGGACAGACAAACCAAAAGTGACGCTTCCTTTGCGA 1872
```

RESULT 7

```
US-10-421-175-1
; Sequence 1, Application US/10421175
; Publication No. US20030219414A1
; GENERAL INFORMATION:
; APPLICANT: Podsakoff, Gregory
; APPLICANT: Watson, Gordon
; APPLICANT: Couto, Linda B.
; APPLICANT: Yang, Bin
; TITLE OF INVENTION: RECOMBINANT ADENO-ASSOCIATED VIRUS VIRIONS FOR THE
; TITLE OF INVENTION: TREATMENT OF LYOSOMAL DISORDERS
; FILE REFERENCE: 0800-0021
; CURRENT APPLICATION NUMBER: US/10/421,175
; CURRENT FILING DATE: 2003-04-22
; PRIOR APPLICATION NUMBER: US/09/715,858
; PRIOR FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 1
; LENGTH: 1956
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1956)
us-10-421-175-1
```

Query Match 10.2%; Score 172.4; DB 17; Length 1956;
Best Local Similarity 52.1%; Pred. No. 3.1e-44;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

```
QY 737 TGAAGCTTGAAGGAGAGTACACTGTGACATCGGAATCAGAACGATCAGTGGAGC 796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 938 TGGGGCTGTGTGACTTCTACACACTCCCTGTGGGATCCGACTGTGCTGCACCA 997
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 797 AGAAGAGCTCTATCTGAACGGGAACCTGTCTTTTGAAGGCTTTGGAAGCAGCAGG 856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 998 AGAGCCAGTTCCTCATCATGGAACCTTTCTATTCCACGCTGTCAACAAGCATGAGG 1057
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 857 AATCCCCGTTCTGGGGCAGGGCACCCTTTATCCATTGATGATAAAGACTTCAACTTC 916
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1058 ATGCGGACATCCGAGGGAAGGGCTTGCAGCTGGCCGCTGCTGTGAAGGACTTCAACTGC 1117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 917 TGAAGTGATCAACGCGAATTCTTTCAGGACCTCTCACTATCCTTACAGTGAAGAGTGGC 976
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1118 TTGCTGGCTTGGTGCCAACGCTTTCGTACCAGCCACTACCCCTATGACAGAGGAAGTGA 1177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 977 TGGATCTTGGCCGACAGACTCGGAATCTTGTGATAGACGAAGCCCCGACGTTGTATCA 1036
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1178 TGCAGATGTGTGACCCGCTATGGGATGTGTGCATCGATGAGTGTCCCGGCGTGGGCTGG 1237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1037 CAAGGTACCA---CTACAATCCCGAGACTCAGAAGATGACAGAAGACACATAAGAGAA 1093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1238 CGCTGCCGAGTTCTTCAACAACGTTTCTTGCAATCACCATGCAAGTGAAGAAAG 1297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1094 TGATGACAGACACAAGAACCATCCAGTGTGATCATGTGGAGTGGCGAAGCAACAG 1153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1298 TGGTGGCTAAGGACAAAGAACCAACCCCGCGGTGATGTGTCTGTGGCCAAACGAGCTG 1357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1154 AGTCCAACCATCCAGACGCGGAGGGTTTCTCAAGCCCTTATGAGACTGCCAATGAAA 1213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1358 CGTCCCACTTAGAATCTGTGCTACTACTGGAAGTGTGATCGCTCACACCAAAATCCT 1417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1214 TGGATCGAACACGCCCCGTTGTGATGAGTGAAGTGAACGACGACAGAGAGAACAA 1273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1418 TGAACCCCTCCCGGCTGTGACCTTGTGAGCAACTTAATATGAG----- 1465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1274 GAGACGTGGCGCTGAAGTACTTGCAGATCGTGTGTGAACAGGTACTACGGCTGTACA 1333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1466 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTGAACAGCTACTCTTGTATC 1525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1334 TCTATCAGGGAAGATAGAAGAGACTTCAAGCTTGAAAAAGACATAGAAGACTCT 1393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1526 ACGACTACGGGCACTGAGTGTGATTCAGCTGCAGCTGGCCACCGATTGAGAACTGGT 1585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1394 ATGCAAGGCACAGAAAGCCCATCTTGTGCACAGAAATTCGTCGCGACGCGATAGTGGCA 1453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1586 ATAGAAGATACAGAAAGCCCATTAATTCAGAGCGAGTATGAGCAGAAAGATTGAGGGGT 1645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1454 TCCACTACGATCCACTCAAAATGTCTCCGAAGATACCAAGCAGAGCTCGTTGAAAAG- 1512
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1646 TTCAACAGATCCACTCTGATGTTCAGTGAAGAGTACAGAAAAAGTCTGCTAGACAGT 1705
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1513 ACGATCAGGCTCCTTTGAAAAA-----AGACTACATCATCGAACAACAGTGGGCT 1567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1706 ACCATCTGGGTCTGATCAAAAAACGAGAAAAATATGTGTTGAGAGCTCATTTGGAATT 1765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1568 TTGCAGATTTTAAGACTCCTCAGAAATGTGAGAAGACCATTTCTCAACCAACAAGGCTTT 1627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1766 TTGCCGATTTTCATGACTGAACAGTCAACGACGAGAGTGTGGGAATAAAAAGGGGATCT 1825
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 1628 TCACAGAGACAGACAAACCCAACTGTTGCTCATGTACTGAGA 1671
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1826 TCACTCGGACAGAGACAAACAAAAGTGCAGCGTTCTTTTGGCA 1869
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 8

```
us-10-388-360-343
; Sequence 343, Application US/10388360
; Publication No. US20030225528A1
; GENERAL INFORMATION:
; APPLICANT: GENOMIC HEALTH
; APPLICANT: Baker, Joffre B.
; APPLICANT: Cronin, Maureen T.
; APPLICANT: Kiefer, Michael C.
; APPLICANT: Shak, Steve
; APPLICANT: Walker, Michael Graham
; TITLE OF INVENTION: GENE EXPRESSION PROFILING IN BIOPSIED TUMOR TISSUES
; FILE REFERENCE: 39740-0001US
; CURRENT APPLICATION NUMBER: US/10/388,360
; PRIOR FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: US 60/412,049
; PRIOR FILING DATE: 2002-09-18
; PRIOR FILING DATE: 2002-03-13
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 343
; LENGTH: 2191
; TYPE: DNA
; ORGANISM: Homo sapiens
us-10-388-360-343
```

Query Match 10.2%; Score 172.4; DB 17; Length 2191;
Best Local Similarity 52.1%; Pred. No. 3.3e-44;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

```
QY 737 TGAAGCTTGAAGGAGAGTACACTGTGACATCGGAATCAGAACGATCAGTGGAGC 796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 964 TGGGGCTGTGTGACTTCTACACACTCCCTGTGGGATCCGACTGTGCTGCACCA 1023
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 797 AGAAGAGCTCTATCTGAACGGGAACCTGTCTTTTGAAGGCTTTGGAAGCAGAG 856
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1024 AGAGCCAGTTCCTCATCATATGGAACCTTCTATTTCCACGGTGTCAACAAGATGAGG 1083
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 857 AATCCCCGTTCTGGGGCAGGGCACCCTTTATCCATTGATGATAAAGACTTCAACTTC 916
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1084 ATGCGGACATCCGAGGGAAGGGCTTTCGACTGGCCGCTGCTGTGAAGGACTTCAACTGC 1143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 917 TGAAGTGATCAACGGAATTCTTTCAGGACCTCTCACTATCCTTACAGTGAAGGTGGC 976
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1144 TTGCTGGCTTGGTGCCAACGCTTTCGTACCAAGCCTATACCGAGGAAGTGA 1203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 977 TGGATCTTGGCCGACAGACTCGGAATCTTGTGATAGACGAAGCCCCGACGTTGGTATCA 1036
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1204 TGCAGATGTGTGACCGCTATGGGATGTGTGATTCAGATGAGTGTCCGGCGTGGGCTGG 1263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1037 CAAGGTACCA---CTACAATCCGAGACTCAGAAAGATAGCAGAAAGCAACATMAAGAA 1093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1264 CGCTGCCGAGTTCTTCAACAACGTTTCTTCGATTCACCATGCAAGTGATGAAGAAAG 1323
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1094 TGATCGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGGAACGAACAG 1153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1324 TGGTGGTGAAGGACAAAGAACCAACCCCGGCTGTGATGTGTGTGTGGCCAAAGCCTG 1383
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1154 AGTCCAACCATCCAGACGCGGAGGGTTCTTCAAGCCCTTTATGAGACTGCCAATGAAA 1213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1384 CGTCCACCTTAGAATCTGTGCTGCTACTACTTGAAGATGTGTGATGCTCACACCAAAATCCT 1443
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1214 TGGATCGAACACGCCCCGTTGTTCATGTGAGCATGATGACGCAACGACGAGAGAACAA 1273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1444 TGAACCCCTCCGGCTGTGACCTTTGTGAGCAACTTAATATGACAG----- 1491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1274 GAGACGTGGCGCTGAAGTACTTGCACATCGTCTGTGTGAACAGTACTACGGCTGTACA 1333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```
Db 1492 CAGACAAAGGGGCTCCGTATGTGATGTGATCTGTTTGAACAGCTACTACTCTTGTATC 1551
OY 1334 TCTATCAGGGAAGATAGAAGAGACTTCAAGCTCTGAAAAAGACATAGAAGACTCT 1393
Db 1552 ACGACTACGGGCACTGGAGTTGATTCAGCTGCAGCTGGCCACCCAGTTTGAGACTGGT 1611
OY 1394 ATGCAAGGCACAGAAAGCCCATCTTTGTCTACAGAAATTCGGTCCGACCGCATAGCTGGCA 1453
Db 1612 ATAAAGATATCAGAAAGCCCATTTATTCAGAGCGAGTATGAGCAGAAACGATTGAGGGT 1671
OY 1454 TCCACTACGATCCACCTCAAAATGTTCTCCGAAGATACCAAGCAGAGCTCGTTGAAAAG- 1512
Db 1672 TTCAACAGATCCACTCTGATGTTCACTGAAGAGTACAGAAAAAGTCTGTAGACAGT 1731
OY 1513 ACGATCAGGCTCCTTTGAAAAA-----AGACTACATCATCGAACAACAGTGTGGGCT 1567
Db 1732 ACCATCTGGGTCTGGATCAAAAACGAGAAAAATATGTGTTGAGAGCTCATTTGGAATT 1791
OY 1568 TTGCAGATTTTAAGACTCCTCAGAAATGTGAGAAGACCCATTCTCAACCAAGGGTGT 1627
Db 1792 TTGCCGATTTTCATGACTGAACAGTCAACGACGAGAGTGTGGGAATAAAAAGGGATCT 1851
OY 1628 TCACAAGAGACAGAACCCAAACTCGTTGCTCATGTACTGAGA 1671
Db 1852 TCACTCGGACAGACAAACCAAAAGTGACGCTTCTTTTGGCA 1895
```

RESULT 9

```
US-10-429-802-16
; Sequence 16, Application US/10429802
; Publication No. US20030228285A1
; GENERAL INFORMATION:
; APPLICANT: HUNG, MIEN-CHIE
; APPLICANT: WONG, KA YIN
; APPLICANT: ZOU, YIYU
; TITLE OF INVENTION: BIPARTITE T-CELL FACTOR (TCF)-RESPONSIVE PROMOTER
; FILE REFERENCE: UTSC:752US
; CURRENT APPLICATION NUMBER: US/10/429,802
; PRIOR FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: 60/377,672
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 16
; LENGTH: 2191
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-429-802-16
```

```
Query Match 10.2%; Score 172.4; DB 17; Length 2191;
Best Local Similarity 52.1%; Pred. No. 3.3e-44;
Matches 492; Conservative 0; Mismatches 431; Indels 21; Gaps 4;

OY 737 TGAACCTGAAAAAGAGTACACTGTGACATCGAATCAGAACGATCAGCTGGAGC 796
Db 964 TGGGGCTGTGTCTGACTTCTACACACTCCCTGTGGGATCCGCACTGTGCTGCACCA 1023
OY 797 AGAAGAGGCTCTATCTGAACGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAGAGG 856
Db 1024 AGAGCCAGTTCTCTCATCAATGGAAACCTTTCTATTCCACGCTGTCACAAAGCATGAGG 1083
OY 857 AATCCCGCTTCTGGGGCAGGGCACCCTTTTATCAATGATGATAAAGACTTCAACCTTC 916
Db 1084 ATGGGACATCGAAGGAAAGGCTTGCACTGCCGCTGCTGTGAAGAACTTCAACCTGC 1143
OY 917 TGAAGTGATCAACGGGAATTTTTCAGGACCTTCACATATCTTACAGTGAAGAGTGGC 976
Db 1144 TTGGCTGGCTTGTGTGCCAACGCTTTCGTAACCACTAACCCCTATGACAGAGAGTGA 1203
OY 977 TGGATCTTCCGACAGACTCGGAATCTTGTGATAGAGGAAGCCCGCAGAGTTGTTATCA 1036
Db 1204 TGCAGATGTGTGACCGCTATGTGATGTGTGTCATCATGAGTGTGTCCCGGCGTGGGCTGG 1263
```

```
OY 1037 CAAGGTACCA---CTACAATCCGAGACTCAGAAAGATAGCAGAAACAAATAGAAGAA 1093
Db 1264 CGCTGCCGAGTTCTTCAACAACGTTTCTGTGCATCACCAACATGCAGGTGATGAAGAA 1323
OY 1094 TGATCGACAGACAAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGGAACCAACAG 1153
Db 1324 TGGTGCGTAGGGAACAAGAACCAACCCCGCGGTGTGATGTGTGTGGCCAAAGACCTG 1383
OY 1154 AGTCCAACCATCCAGACGCGGAGGTTTCTTCAAAAGCCCTTTATGAGACTGCCAATGAAA 1213
Db 1384 CGTCCACCTAAGAAATCTGTGCTACTACTTGAAGATGGTGAATGCTCACACCAAAATCCT 1443
OY 1214 TGGATCGAAACGCCCCGTTGTCTATGTGATGAGCATGATGACCGCACAGACGAGAA 1273
Db 1444 TGAACCCCTCCGCGCTGTGACCTTTGTGAGCACTTAACATATGACG----- 1491
OY 1274 GAGACGTGGCGCTGAAGTACTTGCACATCGTCTGTGTAACAGGTACTACGGCTGTACA 1333
Db 1492 CAGACAAAGGGGCTCCGTATGTGATGTGATGTGATGTTTGAACAGCTACTCTTGTGATC 1551
OY 1334 TCTATCAGGGAAGATAGAAGAGACTTCAAGCTCTGAAAAAGACATAGAAGACTCT 1393
Db 1552 ACGACTACGGGCACTGAGTTGATTCAGCTGCAGCTGGCCACCCAGTTTGAGAACTGGT 1611
OY 1394 ATGCAAGCAGAGAAAGCCCACTTTGTCTCAGAAATTCGGTGGGAGCGGATAGCTGCA 1453
Db 1612 ATAAAGATATCAGAAAGCCCAATTAATCAGAGCAGTATGAGCAGAAACGATTGACGGT 1671
OY 1454 TCCACTACGATCCACCTCAAAATGTTCTCCGAAGATACCAGAGAGCTCGTTGAAAAAG- 1512
Db 1672 TTCACCAAGATCCACTCTGATGTTCACTGAGAGATACAGAAAAAGTCTGCTAGAGCAGT 1731
OY 1513 ACGATCAGGCTCCTTTGAAAAA-----AGACTACATCATCGAACAACAGCTGTGGGCT 1567
Db 1732 ACCATCTGGGTCTGGATCAAAAACGAGAAATATGTGTTGAGAGACTCATTTGGAATT 1791
OY 1568 TTGCAGATTTTAAGACTCCTCAGAAATGTGAGAAGAACCCATTCTCAACCAAGGCTGTT 1627
Db 1792 TTGCCGATTTTCATGACTGAACAGTCAACGACGAGAGTGTGGGAATAAAAAGGGATCT 1851
OY 1628 TCACAAGAGACAGAACCCAAACTCGTTGCTCATGTACTGAGA 1671
Db 1852 TCACTCGGACAGACAAACCAAAAGTGACGCTTCTTTTGGCA 1895
```

RESULT 10

```
US-10-430-503-7
; Sequence 7, Application US/10430503
; Publication No. US20040005684A1
; GENERAL INFORMATION:
; APPLICANT: HUNG, MIEN-CHIE
; APPLICANT: LAN, KENG-LI
; APPLICANT: OU-YANG, FU
; APPLICANT: LIU, JAW-CHING
; APPLICANT: LAN, KENG-HSIN
; TITLE OF INVENTION: TARGETING PROTEINS TO DELIVER THERAPEUTIC OR DIAGNOSTIC
; FILE REFERENCE: UTSC:797US
; CURRENT APPLICATION NUMBER: US/10/430,503
; PRIOR FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: 60/383,063
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 7
; LENGTH: 2191
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-430-503-7
```

```
Query Match 10.2%; Score 172.4; DB 17; Length 2191;
Best Local Similarity 52.1%; Pred. No. 3.3e-44;
```

	Matches	492;	Conservative	0;	Mismatches	431;	Indels	21;	Gaps	4;
OY	737	TGGAACTT	GA	AAAAAG	ACGATAC	CTCTG	GACATCGGA	ATCAGAA	CGATCAGCTGG	ACG 796
Db	964	TGGGGCCT	GTGTCT	GACTTCT	TACACACT	CCCTGT	GGGATCGCA	CTGTGCT	GTCA	CCA 1023
OY	797	AGAAGAGC	CTCTA	TCTGA	ACGGGAA	CCCTGT	CTTTTGA	GGGCTTTG	AAAGCAG	CAGG 856
Db	1024	AGAGCCAG	TTCTCT	ATCAT	TGGGAA	CCTTTCTA	TTTCCAC	GGGTCTCA	AGCATGAG	1083
OY	857	AAATCCCC	GGTTC	TGGGGC	AGGGCA	CTTTAT	TCATGT	ATGATTA	AAAGACTT	CAACCTTC 916
Db	1084	ATTCGGGA	CATCC	AGGGAA	GGGCTTC	GACTGG	CCGCTGT	GTGAAG	GAATTCA	ACCTGC 1143
OY	917	TGAAGTGA	TCAAC	CGGAAT	TTCTTT	CAGAC	CTCTCA	TATCCTT	ACAGTGA	AGATGGC 976
Db	1144	TTCCGTGG	CTTGG	TGCCA	ACGCTTT	CCGTAC	CAACCA	CTACCTT	ATGCA	AGAAAGTGA 1203
OY	977	TGGATCTT	GCCGAC	AGACTCG	GAATCCT	TGTGT	ATAGAC	GAAAGCCCCG	CACGTTG	TATCA 1036
Db	1204	TGCAGATG	TGTGA	CCGCTAT	GGATTT	GTGTCA	TGATGAT	GTCTCCG	CGTGGG	CTGG 1263
OY	1037	CAAGGTAC	CA--	CTACAT	TC	CGAGACT	CAAGATAG	CAGAA	CAACAT	TAAGAA 1093
Db	1264	CGGTCCG	CAGTTCT	TCAACAA	CGTTTCT	CTGTCA	TACCA	CATGCAG	TGATGA	AGAA 1323
OY	1094	TGATCGAC	ACAGACA	CAAGAAC	CCATCC	AGTGT	ATCATGT	GAGTGTG	CGAAGAA	CCAG 1153
Db	1324	TGTGCGTA	GGGACA	CAAGAAC	CCCCGG	GTGTGAT	GTGTGTG	GGCCA	AGAC	CTG 1383
OY	1154	AGTCCAA	CCATTC	CAGACG	CGGAG	GGTTCT	TCAAAG	CCCTTTAT	TAGAC	CTGCCAATGAAA 1213
Db	1384	CGTCCAC	CACTAGA	ATCTGT	GCTACT	TA	GAAGATG	TGATGCT	CAACCA	AAATCT 1443
OY	1214	TGGATCGA	ACACG	CCCCCG	TGTGT	CATGTG	TAGCAT	GTGACG	CAACGCA	CGAGAA 1273
Db	1444	TGGACCC	CTCCCG	CCCTGT	GACTTT	GTGAG	CAACTT	MACTAT	GACAG--	----- 1491
OY	1274	GAGACGTG	CGCTGA	AGTACTT	GCACAT	CGTCT	GTGTGA	CAGSTACT	ACGCTG	TACA 1333
Db	1492	CAGACA	AGGGGG	CTCCGT	ATGTGAT	GTGATCT	GTTTGAA	CAGCTACT	ACTCTT	GTAATC 1551
OY	1334	TCTATCAG	GGGAAG	ATAGA	AGGA	CTTCA	AGCTCT	GAAAAA	AGACAT	AGAAGCTCT 1393
Db	1552	ACGACTAC	GGGCACT	GAGTTG	ATTCAG	CTGCA	GCTGG	CCACCAG	TTTGA	AACTGGT 1611
OY	1394	ATGCAAGG	CA	CAGAAAG	CCCCAT	CTTTGT	CA	CAGAA	TTCCG	TGGGACGCGATGCTGGCA 1453
Db	1612	ATTAAGAA	GTATCAG	AAAGCC	CAATTA	TTCA	GAGCAG	ATATGAG	CAAGAA	ACGATTCGACGGGT 1671
OY	1454	TCCACTAG	ATCCA	CCCTCA	AAATGTT	TCTCC	GAAAGT	ATCCAA	GACAGAGCT	CGTTGAAAG- 1512
Db	1672	TTTCA	CCAGAT	CACTCT	GTATGT	TTCACT	GAAGAT	ACCAGAA	AAAGTCTG	CTAGAGCAGT 1731
OY	1513	ACGATCAG	GGCTCT	TTTGA	AAAA-----	AGACTA	CATCAT	CGGA	ACACA	CGTGTGGGCT 1567
Db	1732	ACCATCTG	GGCTG	ATCA	AAAAAC	GCAGAA	AAATAT	GTGTTG	GAAGCTCA	TTTGAAT 1791
OY	1568	TTGCAGAT	TTTTA	AGACTCT	CTCAG	AATGTG	AGAAGAC	CCATTT	CAACCA	CAAGGGTGT 1627
Db	1792	TTGCCGAT	TTTCA	TGACTG	AACTGT	CA	CCGACG	AGTGTG	GGGAAT	MAAAAGGGATCT 1851
OY	1628	TCACAAG	AGACAG	CAACCC	CAAACT	CGTTG	CTCATGT	ACTGAG	A 1671	
Db	1852	TCACTCGG	CAGAGCA	CAACCA	AAAAAG	TGCAG	CGCTTCT	TTTGG	CA 1895	

RESULT 11
US-10-335-053-51
; Sequence 51, Application US/10335053
; Publication No. US20040241653A1
; GENERAL INFORMATION:
; APPLICANT: Quark Biotech, Inc.

```

; TITLE OF INVENTION: Methods for identifying marker genes for cancer
; FILE REFERENCE: 68733-A; 070/US1
; CURRENT APPLICATION NUMBER: US/10/335,053
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: 60/345,317
; PRIOR FILING DATE: 2001-12-31
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 51
; LENGTH: 2191
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-335-053-51

```

Query Match	10.2%	Score 172.4	DB 20	Length 2191
Best Local Similarity	52.1%	Pred. No. 3.3e-44		
Matches 492	Conservative	0	Mismatches 431	Indels 21
				Gaps 4

QY	737	TGGAACCTTGAAAAAGACGAGTA	CACTCTGGACATCGGAATCAGAACGATCAGCTGGGACG	796
Db	964	TGGGGCCTGTGTCTGACTTCTA	CACACTCCCTGTGGGATCCGCAC	1023
QY	797	AGAAAGGCTCTATCTGAA	CGGGAAACCTGTCTTTTGAAGGGCTTTGGAAACACGAGG	856
Db	1024	AGAGCCAGTTCCTCATCA	ATGGGAAACCTTCTATTTCACGGGTGTCAACAAAGCATGAGG	1083
QY	857	AATTCCTCGTCTGGGGCAGGG	CACCTTTTATTCATTGTATGATTAAGACTTCAACCTTC	916
Db	1084	ATCGGACATCCGAGGGAA	GGGCTTCGACTGGCCGCTGTGTGAAGGACTTCAACCTGC	1143
QY	917	TGAAGTGATTCAA	CGCGAATCTTTCAGGACCTCTCACTATCTTCA	976
Db	1144	TTCCGCTGGCTTG	GTGTCACCGCTTTCCGTACCA	1203
QY	977	TGGATCTTGCCGACAG	ACTCGGAATCCTTGTCATAGACGAAGCCCGCACCTTGCTATCA	1036
Db	1204	TGCAGATGTGTAC	CGCTATGGGATTTGTGTATCGATGAGTGTCCCGGCGTGGGCTGG	1263
QY	1037	CAAGGTACCA---	CTACAATCCCGAGACTCGAAGATAGCAGAAAGACAACTAAGAGAA	1093
Db	1264	CGGTCCCGCAGTTC	TCAACAACGTTTCTCTGCATCACCAATGCAGGTGATGGAAGAG	1323
QY	1094	TGATCGACAGACACA	GAAGACCATCCCACTGTGATCATGTGAGCTGTGGCAACGAACAG	1153
Db	1324	TGTTGCGTAGGAGACA	GAAGACCAACCCCGCGTGTGATGTGTGTGTGTCACAGAGCCTG	1383
QY	1154	AGTCCAACCATCC	AGACGCGGAGGGTTCCTCAAGCCCTTATGAGACTGCAATGAAA	1213
Db	1384	CGTCCACCTAGAAT	CTGTGTGCTACTACTTGAAGATGTGTATCGCTCACACCAATCCT	1443
QY	1214	TGGATCGAACA	CGCCCCGTTGTCTATGTGACATGATGACCGACACGACGAGAACAA	1273
Db	1444	TGGAACCCCTCC	CGGCGCTGTGACCTTGTGAGCACTTA	1491
QY	1274	GAGACGTGGCGCT	GAACTTTCGACATCTGTGTGTGAACAGTACTACGGCTGTACA	1333
Db	1492	CAGACAAGGGGCT	CCGTATGTGATGTGATCTGTTTGAACAGCTACTACTTGGTATC	1551
QY	1334	TCTATCAGGGA	AGATAGAAAGAACTTCAAGCTCTGAAAAAAGACATAGAAAGCTCT	1393
Db	1552	ACGACTACGGGCA	CTGAGTTGATTACGTGCACTGGCCACCCAGTTTGAAGAACTGGT	1611
QY	1394	ATGCAAGGCACA	GAAGCCCACTTTGTGCAGAAATTCGGTGGCAGCGAATAGCTGCA	1453
Db	1612	ATAAGAAGTAT	CAGAAAGCCATTTATTCAGAGCGAGTATGAGCAGAAAGATTTGACAGGT	1671
QY	1454	TCCACTACGAT	CCACCTCAATGTCTCCGAAGATACCAAGCAGAGCTCGTTGAAAAG-	1512
Db	1672	TTTCAACGAGAT	CCACTCTGTGATCTGAAAGAGTACCAAGAAAGTCTGCTAGAGCAGT	1731
QY	1513	ACGATCAGGCT	CTTTTGAAAAA-----AGA	1567
Db	1732	ACCATCTGGGT	CTGATCAAAAAACGACAGAAATATGTGTGTGAGAGCTCATTTGGAATT	1791


```
QY 919 AAGTGGATCAACGCGAATTCTTTCAGGACCTCTCACTATCCTTACAGTGAAGAGTGGCTG 978
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1108 CGTTGGCTCTGGGGCAAAATTCCTTTCGTACCAAGCCACTATCCCTACTCAGAGAGGACTT 1167
QY 979 GATCTTGGCCGACAGACTCGGAATCCTTGTGTATAGACGAAGCCCCGACGTTGG--TATC 1035
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1168 CAGCTCTGTGACCGATACGGGATTTGTGTCTCATCATGATGAGTGTCCCGGTGTGGCATTTGTG 1227
QY 1036 ACAAGGTACCACTACATCCCGAGACTCAGAAGATAGCAGAAGACAATAGAAGAAATG 1095
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1228 CTACCTCAGAGTTTGGCAACGAGTCACTTCGGCACCACTTAGAGGTGATGGAGAGCTG 1287
QY 1096 ATCGACAGACACAAGAACCATCCAGTGTGATCATGTGAGTGTGGCGGAACGAACAGAG 1155
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1288 GTTCGCGCGGACAAAATCAACCTGCGGTGTGATGTGTCTGTGGCCAATGAGCCTTCC 1347
QY 1156 TCCAACCATCCAGACGCGGAGGGTTCTTCAAGCCCTTTATGAGACTGCCAATGAATG 1215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1348 TCTGCTCTGAAACCGCGCGCATATTACTTTAAGACGCTGATCACCAACACCAAGCCCTG 1407
QY 1216 GATCGAACAACGCCCCGTTGTCTATGTGAGCATGATGAGCGCACAGACGAGAGAACAAGA 1275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1408 GACCTCAACCCGTCGCGTGACCTTTGTGAGCA-----ACGCCAATATGATGCA 1455
QY 1276 GACGTGGCGCTGAAGTACTTTCAGACATCGTCTGTGTGAACAGGTACTACGGCTGTACATC 1335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1456 GACCTGGGGGCCCCGTAACGTGATGTATCTGTGTAAACAGTACTTTCTTGTGTATCAT 1515
QY 1336 TATCAGGGAAGATAGAGAAGGACTTCAAGCTCTGGAAGAAAGACATAGAAGAGCTTAT 1395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1516 GACTATGGGCATTGGAGGTGATTCAGCCACAGCTGAATAGCCAGTTTGAGAACTGTAT 1575
QY 1396 GCAAGGCAACAGAAAGCCCATCTTGTGTCACAGAAATCGGTGCGGACGCGATAGCTGCATC 1455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1576 AAGACGCATCAGAAAGCGATTATCAGAGCGAGTATGAGACAGACGCAATCCAGGAGATC 1635
QY 1456 CACTACGATCCACCTCAATGTCTTCGGAAGAGTACCAAGCAGAGCTCGTTGAAGAACG 1515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1636 CACAGGAGACCCGCTCGCATGTTCAGTGAGAGTACCAAGAGCTGTTCTGAGAAATTAC 1695
QY 1516 ATCAGGCTCTTTTGA---AAAGACTACATCATCGGAACACAGTGTGGGCTTT 1569
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1696 CATTCAGTTCTGATCAGAAACGTAAAGATACGTGTCGAGAGCTCATCTGGAATTTTC 1755
QY 1570 GCAGATTTTAAGACTCTCAGAAATGAGAGAGCCATTCTCAACACAAAGGTTTTC 1629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1756 GCGCACTTCATGACGAACAGTCAACGCTGAGAGTATCGGAACAAAGAGGATCTTC 1815
QY 1630 ACAAGAGACAGAACCCAACTCGTTGCTCATGTACTGAGA 1671
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1816 ACTCGCCAGAGACAGCCCAAACTTCGGCCTTATTTTGGCA 1857
```

RESULT 14

```
US-10-757-093-3
; Sequence 3, Application US/10757093
; Publication No. US20050153448A1
; GENERAL INFORMATION:
; APPLICANT: CAMBIA
; TITLE OF INVENTION: Fungal beta-glucuronidase genes and gene products
; FILE REFERENCE: 415
; CURRENT APPLICATION NUMBER: US/10/757,093
; CURRENT FILING DATE: 2004-01-14
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 1905
; TYPE: DNA
; ORGANISM: Penicillium canescens isolate RPK
US-10-757-093-3
```

Query Match 8.4%; Score 141.6; DB 22; Length 1905;
Best Local Similarity 49.1%; Pred. No. 3e-34;

```
Matches 485; Conservative 0; Mismatches 449; Indels 54; Gaps 2;
QY 744 TGAAGAGACGAGTACACTGTGACATCGGAATCAGAACGATCAGCTGGAGCAGAGAG 803
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 906 TGTAGTCACACCTACAAATTTGGCTACGGGCGTGGCTACTGTCAAGGTGCGGGTCA 965
QY 804 GCTCTATCTGAACGGGAAACCTGTCTTTTGAAGGCTTTGAAAGCAAGAGAAATCCC 863
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 966 ATTCTTAATTAATGAAAGCCCTTCTACTTACCTTACCGGTTTGGCAACATGAAGACACAGC 1025
QY 864 CGTTCTGGGCGAGGCACTTTTATCATTGATGATTAAGACTTCAACCTTCTGAAGTG 923
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1026 AGTACGTGGCAAGGACATGACCCAGCATACATGCTTACGATTTCCAACCTCATGAATG 1085
QY 924 GATCAACGCGAATCTTTCAGACCTCTCACTATCCTTACAGTGAAGAGTGGCTGATCT 983
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1086 GATTGAGCAAAATCTTTTCGGACTTCACTATCTTACGCGGAAGAGTCATGATTT 1145
QY 984 TGCCGACAGACTCGGAATCCTTGTGATGACGAAGCCCCGACAGTTGGT----- 1032
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1146 CGCAGATCGAAATGGAATTTGTCTGATGATGAACAACCTGCCGTTGGTCTGAACATTGC 1205
QY 1033 -----ATCACAAGGTACCACTACAATCC 1055
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1206 CTTGATGGCGGTATCTGAGAGTGTGCCCAACAACATTTACGCCAGATGCGATTAACGA 1265
QY 1056 CGAGACTCAGAAAGATAGCAGAAGACAACATAGAAGATGATGACAGACACAAGAACCA 1115
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1266 TAAACCCAAAGAGCCCAACAGCAGGCGATTCTGTAGCTCATTTGCCAGACAAAACCA 1325
QY 1116 TCCAGTGTATCATGTGAGTGTGGCGAAGCAAGACAGAGTCCAACCATCCAGACGCGGA 1175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1326 TGCCAGTGTGTATGTGTGTCTATTTGCCAAGAGCCGCTCATCTGAAGATGAGCTCG 1385
QY 1176 GGGTTCTTCAAGCCCTTATGAGACTGCCAATGAATGATGCAACAGCCCGTTGT 1235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1386 CGAATACTTCAGGCACTGACCAATTTGACTCTGCAACTGATCCAACCTGCCCTATTAC 1445
QY 1236 CATGTGAGCATGATGAGCAGCAGACAGAGAGAACAGAGACGTGGCGCTGAAGTACTT 1295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1446 ATTTGCTAACGTGGCAACGCGACATATCAGCTGTGATCGGATCTGTATCTG-----TT 1499
QY 1296 CGACATCGTCTGTGTAACAGTACTACGCGTGTATCATCTATCAGGAGAGATAGAAGA 1355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1500 TGAATGTCAGTTGCAATAATCGGTATTTCCGATGTATTTCTCAACAGAGAACCTTGAAGA 1559
QY 1356 AGCACTTCAAGCTCTGGAAGAAAGACATAGAAGAGCTCTATGCAAGGCAACAGAACCCCAT 1415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1560 AGCAGAGGACAGCTCTTGAAGAGAGCTGCATGATGGCAAGAGAAATTCCACAGGCCGAT 1619
QY 1416 CTTTGTACAGAAATTCGGTGCAGACGGATAGCTGGCATCCACTACGATCCACTCAAT 1475
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1620 CGTCATGACCGAATATGTTGTCAGATTAACCTTGCAGGCTTCACTCTATCTCGGACTGCC 1679
QY 1476 GTTCTCCGAAGTACCAAGCAGAGCTCGTTGAAAAAGAGATCAGGCTCCTTTGAAAAA 1535
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1680 TTGAGCGGAAGAGTTCCAAGTACAAATGCTAGACATGTACCATGAGTGTGATCGCAT 1739
QY 1536 AGACTACATCATCGGAACACACAGTGTGGGCTTTGCAAGTTTAAAGACTCCTCAGAAATG 1595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1740 TGAATGATGGCAGGCGGAGCATGTTTGAACCTTCGCGATTTCAGAACCAACTTGGGTAT 1799
QY 1596 GAGAGAGCCCATTTCTCAACCAAGAGGTGTTTTCAGAGAGACAGACCAACCAACTCGT 1655
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1800 CATCCGAGTAGACGGTACCAAGAGGGGTGTTTCAACCCGTGACCGAAGAACCAAGGCGGC 1859
QY 1656 TGCTCATGTACTGAGAGAGACTGTGAGT 1683
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1860 AGCTCATAGTTGAGGGGCAAGGTGACT 1887
```

RESULT 15

US-10-120-145-7

```
; Sequence 7, Application US/10120145
; Publication No. US20030157684A1
; GENERAL INFORMATION:
; APPLICANT: Jefferson, Richard A.
; APPLICANT: Kilian, Andrzej
; APPLICANT: Keese, Paul Konrad
; TITLE OF INVENTION: MICROBIAL BETA-GLUCURONIDASE GENES, GENE PRODUCTS AND
; FILE REFERENCE: 190106.405
; CURRENT APPLICATION NUMBER: US/10/120,145
; PRIOR FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/149,727
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/058,263
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-09
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1887
; TYPE: DNA
; ORGANISM: Bacillus sp.
; US-10-120-145-7
```

```
Query Match      8 4%; Score 141.4; DB 16; Length 1887;
Best Local Similarity 45.4%; Pred. No. 3.5e-34;
Matches 734; Conservative 0; Mismatches 822; Indels 60; Gaps 4;
```

```
QY 120 CCAGATCTGTGCTCGAAGAAGACCTTCACTCAAAACCACTTCTACGTTCCGAA 179
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 208 CAAGAAATCCGCAACCATATCGATATGCTGTGTAAGAACGTGAGTTACGGTCCGGC 267

QY 180 GNACTTTCACAAAACACATCAGACTTACTTTGCTGCGGTGAACAAGACTGGAGGT 239
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 268 CTATCTGAAGATCAGCGTATCGTCTCGCTCGCTCTGCACTCAAAAGCAATTGT 327

QY 240 CTTCTCAACGAGAGAAAGTGGAGAGATCACAATTGAATACCTTCCCTTGAAGTAGA 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 328 CTATGTCAATGTGAGCTGTGCTGAGACACAAGGCGGATTCCTGCCATTGGAAGCGGA 387

QY 300 TGTGACGGGGAAGTGAATCCGAGAGAACGAACTCAGGTTGTTGAGAACAGATT 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 388 AATCAACAACCTCGCTGCGTATGAGATCGGTCACCGTCCGCTGGAACAATCCT 447

QY 360 GAAAGTGGAGATTTCCCTCGAAGTTCCAGACAGCGGCACTCAACCGTGGATT 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 448 CGAGATAGCAACCTCCCGGTGGGTGTACAGGAGCGCCACGAAGAGGCGCTCGAAA 507

QY 420 TGAAGTTTTCACCTGCAAACTTGAATTTCCCTTACGTTGAAATCATTAAGCCTGT 479
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 508 AGTCATTCTGAACAGCCGAACTTGAATTTCACTATGACAGGCTGCAACCGTCCGGT 567

QY 480 TCTGATAGATTCAAGACCAACGAGATCTGCAATCTGGGTGGAACGAGTGAATC 539
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 568 GAAATCTACACGACCCCGTTTACGTAAGTGAAGCATCTCGGTTGTGACCGCACTTCAA 627

QY 540 TGAACCGGAGAGAACTTGGAAAAGTGAAGTGAAGTGAAGTCTCAGAAAGCGGT 599
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 628 TGGCCCAACCGGACTGTGACCTATACGTTGAGCTTTCAAGGCAAAAGCCGAGACGTGAA 687

QY 600 GGGACAGAGATGACATCAAACTTGAAGAGAGAGAAAAAGATTAAGACATCCAACG 659
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 688 AGTGTCCGTCGTGATGAGGAAGGCAAGTGTGCGCAAGCAACCGAGGCGCTGAGCGGTAA 747

QY 660 ATTCTCGAAGGGGAGTTCTCTCGAAAAACGCCAGTTCTGAGCCTCGAAGATCCATA 719
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 748 CGTGAGATTTCGAATGTCTCTCTGGGAACCACTGAACACGTAATCTTAACCAATCAA 807

QY 720 TCTTTATCTCTCAAGTGGAACTTGAAGAAAGACGAGTACATCTGACATCGGAATCAG 779
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 808 AGTGAACACTGTGTAACGACGGACTGACATCGATGTCTATGAAGAGCCGTTCCGCGGTGCG 867

QY 780 AACGATCAGCTGGGACGAGAGAGGCTTATCTGAACGGGAACCTGTTTGAAGGG 839
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
DB 868 GACCGTGAAGTCAACGACGCGCAAGTTCTCTCATCAACAACAAACCGTTCTACTTCAAGG 927
QY 840 CTTTGGAAAGACGAGGAATTTCCCGTTCTGGGGGACGGCACCTTTATCCATTGATGAT 899
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 928 CTTTGGCAACATGAGGACACTCTTATCAACGGCCGTTGGCTTTAAACGAAGCAGCAATGT 987
QY 900 AAAAGACTTCAACCTTCTGAAGTGGATCAACCGCAATTTCTTTCAGGACCTCTCATATCC 959
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 988 GATGATTTCATATCTCTCAATATGATCGGCGCCACAGAGCTTCCGAGCCGACATATCC 1047
QY 960 TTACAGTGAAGATGGCTGATCTTGGCCGACAGACTCGGAATCTTGTGATAGACGAAGC 1019
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1048 GTACTCTGAAGATTGATGCTCTTGGCGGATCGGAGGCTTGGTGTGATGACGAGAC 1107
QY 1020 CCGGACGTTGGTAT----- 1034
DB 1108 TCCGGCAGTTGGCGTGCACCTCAACTTCATGGCCACCAACGGGACTCGGCGGAAGCAGCGA 1167
QY 1035 ---CACAGGTACCACTACATCCCGAGACTCAGAAAGATAGCAAGAACATTAAGAG 1091
DB 1168 GCGCGTCAGTACCTGGGAGAGATTTCGGACGTTTGAACCATCAAGACGTTCTCCGTGA 1227
QY 1092 AATGATCGACAGACAAAGAACCATCCCACTGTGATCATGTGAGTGTGGCGAAGCAACC 1151
DB 1228 ACTGTGTCTCTGACAAAGAACCATCAAGCGTCTGTATGTGAGCATCGCCAAACGAGGC 1287
QY 1152 AGAGTCCAAACATCCAGACGCGGAGGTTTCTCAAAAGCCCTTTATGAGACTGCCAATGA 1211
DB 1288 GCGGACTGAGGAAGAGGCGCGGTACGAGTACTTCAAGCCGTTGGTGAAGCTGAACCAAGGA 1347
QY 1212 AATGATCGA---ACAGCCCGTTGTCTATGTGAGATGATGAGACGACCAACGAGAG 1268
DB 1348 ACTGACCCACAGAAAGCGTCCGTCACGATCGTGTGTTGTGATGAGTGAATCCCGGAGAC 1407
QY 1269 AACAAAGACGTGCGGTGAAGTACTTTCGATCTGTGTGTAACAGTACTACGCGT 1328
DB 1408 GGAACAAGTCCCGAACTGA-----TTGAGCTCATCGCGCTCAATCGCTAATAACGAGATG 1461
QY 1329 GTACATCTTACAGGGAAGATGAAGAG---GACTTCAAGCTCTGMAAAAGACATAGA 1385
DB 1462 GTACTTGAATGCGGTGATCTGAAGCGGCCAAAGTCCATCTCCGCCAGAAATTTCAAGC 1521
QY 1386 AGAGCTTATGCAAGGACAGAAAGCCCATCTTTGTCAAGAAATTCGGTGGGACGCGAT 1445
DB 1522 GTGAAACAAGCGTTGCCAGAAAGCCCATATGATCACTGATGATACGGCGACACACCGT 1581
QY 1446 AGCTGCATCACTAGATCCACCTCAATGTCTCCGAAGATACCAAGCAGAGCTCGT 1505
DB 1582 TGGGGCTTTCAAGCATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1641
QY 1506 TGAAGAAGCATCAGGCTCTTTTGAAGAAAGCTATCATCATCGGAACACAGTGTGGGC 1565
DB 1642 CCAGGCGAACCAAGTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1701
QY 1566 CTTTGCAGATTTTAAGCTCTCAGATGAGAAAGACCAATTTCAACCAAGGCGT 1625
DB 1702 CTTGCGGACTTTCGCACTCTCAGAGGCGGTGATGCGGCTCCAAGAAACAAAGAGGCGT 1761
QY 1626 TTTCAACAGACAGACAAACCAACTCTGTGCTCATGTACTGAGAAAGCTGTGA 1681
DB 1762 GTTCACTCGTACCGCAAGCGAAAGCTCGCGCGCACGCTCTTTCGCAAGCGCTGGA 1817
```

Search completed: September 1, 2005, 21:26:50
Job time : 1170 secs